

UNIVERSIDAD AUTÓNOMA DE MADRID
ESCUELA POLITÉCNICA SUPERIOR



Grado en Ingeniería Informática

TRABAJO FIN DE GRADO

Análisis de datos sanitarios aplicando metodología CRISP-DM

Rafael Nishizaki Fernandes
Tutor: Jack Mario Mingo Postiglioni
Ponente: Iván Cantador Gutiérrez

Julio 2017

Análisis de datos sanitarios aplicando metodología CRISP-DM

AUTOR: Rafael Nishizaki Fernandes
TUTOR: Jack Mario Mingo Postiglioni
PONENTE: Iván Cantador Gutiérrez

Dpto. de Ingeniería Informática
Escuela Politécnica Superior
Universidad Autónoma de Madrid
Julio de 2017

Resumen (castellano)

Este Trabajo Fin de Grado consiste en el análisis de los datos sanitarios aplicando la metodología CRISP-DM. Actualmente, el Ministerio de Sanidad, Servicios Sociales e Igualdad, junto al Centro de Investigaciones Sociológicas, realiza anualmente un barómetro sanitario con el objetivo de conocer, a nivel general, la opinión y valoración de los usuarios residentes en España, con respecto a los servicios de la sanidad pública española. Con este Trabajo de Fin de Grado, se pretende ampliar el conocimiento obtenido a partir de estudios estadísticos realizados con los resultados del barómetro sanitario, identificando y descubriendo distintos sectores (clústeres) de la población que agrupan a los usuarios de la sanidad pública, que opinan de manera similar respecto a ella. Una vez encontrados los distintos grupos de usuarios, se pretende analizar la relación entre las respuestas dadas en las encuestas, para encontrar patrones e identificar los factores que influyen en la opinión de los usuarios.

Para la realización de este Trabajo de Fin de Grado, se han seguido las actividades descritas en las fases de la metodología CRISP-DM, considerada como una de las más utilizadas y recomendadas para proyectos de minería de datos.

La metodología CRISP-DM consta de seis fases. La primera pretende conocer los objetivos del negocio, valorar la situación actual y realizar un plan del proyecto. La segunda fase pretende conocer los datos iniciales de los que disponemos para la realización del proyecto. En la tercera fase se preparan los datos recogidos en la fase anterior, para que en la cuarta fase se aplique una técnica de modelado que se adapte a los objetivos del negocio y genere información nueva y relevante. En la quinta fase se analizan los resultados obtenidos valorando su calidad, para que, por último, se indiquen los planes de distribución del proyecto.

Palabras clave: Minería de Datos, Metodología CRISP-DM, Barómetro Sanitario, Centro de Investigaciones Sociológicas (CIS), Clúster, Técnicas de Modelado, Aprendizaje Automático

Abstract (English)

This Bachelor Thesis consists of the analysis of health data using the CRISP-DM methodology. Currently, the Ministry of Health, Social Services and Equality, together with the Center for Sociological Research, annually performs a barometer with the aim of knowing, at a general level, the opinion and valuation of users in Spain, regarding the services of the Public health. The objective of this project is to expand the knowledge obtained from statistical studies performed with the results of the health barometer, identifying and discovering different sectors (clusters) of the population that group public health users, who think about it in a similar way between them. Once it is found the different groups of users who think in a similar way, we are going to try to analyze the relationship between the responses given in the surveys, to find patterns and identify the factors that influence the opinion of users.

For the implementation of this project, we followed the activities described in the phases of the CRISP-DM methodology, considered as one of the most used and recommended for data mining projects.

The CRISP-DM methodology consists of six phases. The first aims to know the objectives of the business, know the actual situation and make a project plan. The second phase aims to know the initial data that we have for the realization of the project. The third phase prepares the data collected in the previous phase, so that in the fourth phase we apply a modeling technique that adapts to the business objectives and find new and relevant information. In the fifth phase, we will analyze the results obtained and assess their quality, so that, finally, we indicate the distribution plans of the project.

Keywords: Data Mining, Methodology, CRISP-DM, Sanitary Barometer, Center for Sociological Research, Cluster, Modeling Techniques, Machine Learning

Agradecimientos

En primer lugar, me gustaría agradecer a mi profesor y tutor Mario, por su paciencia, confianza y por todo el conocimiento que ha compartido conmigo, para que fuera posible la realización del presente proyecto.

También quiero agradecer a Laura, por todo su apoyo y cariño, y por siempre estar a mi lado.

Por último, pero no por ello menos importante, agradecer a mi familia, mis padres, mis hermanas y mis amigos.

INDICE DE CONTENIDOS

| | |
|--|----|
| 1 Introducción | 1 |
| Motivación | 1 |
| Objetivos | 1 |
| Organización de la memoria | 2 |
| 2 Estado del arte..... | 3 |
| Introducción | 3 |
| Minería de datos | 3 |
| 2.1.1 Técnicas de minería de datos..... | 3 |
| Metodología CRISP-DM | 3 |
| 3 Análisis de Datos Sanitarios | 7 |
| Fase de comprensión del negocio..... | 7 |
| 3.1.1 Determinar el objetivo del negocio..... | 7 |
| 3.1.2 Valorar la situación..... | 7 |
| 3.1.3 Determinar los objetivos de la Minería de Datos..... | 9 |
| 3.1.4 Realizar un plan de proyecto..... | 9 |
| Fase de comprensión de datos | 10 |
| 3.1.5 Recolección de datos iniciales..... | 10 |
| 3.1.6 Descripción de los datos..... | 10 |
| 3.1.7 Exploración de los datos | 11 |
| 3.1.8 Verificación de la calidad de los datos..... | 13 |
| Fase de preparación de los datos..... | 13 |
| 3.1.9 Selección de datos..... | 13 |
| 3.1.10 Limpieza de datos | 14 |
| 3.1.11 Estructuración de los datos..... | 14 |
| 3.1.12 Integración de los datos..... | 15 |
| 3.1.13 Formateo de los datos. | 16 |
| Fase de modelado | 16 |
| 3.1.14 Selección de la técnica de modelado | 16 |
| 3.1.15 Generación de un plan de prueba..... | 17 |
| 3.1.16 Construcción del modelo..... | 17 |
| 3.1.16.1 K-Means | 19 |
| 3.1.16.2 Kohonen..... | 22 |
| 3.1.16.3 Bietápico | 24 |
| 3.1.17 Evaluación del modelo. | 26 |
| Fase de Evaluación | 28 |
| 3.1.18 Evaluación de los resultados..... | 29 |
| 3.1.19 Proceso de revisión | 32 |
| 3.1.20 Determinar próximos pasos..... | 32 |
| Fase de Implementación | 32 |
| 3.1.21 Plan de implantación | 33 |
| 3.1.22 Plan de monitoreo y mantención..... | 33 |
| 3.1.23 Informe final..... | 34 |
| 3.1.24 Revisión del proyecto | 34 |
| 4 Conclusiones y trabajo futuro | 35 |
| Conclusiones | 35 |
| Trabajo futuro | 35 |

| | |
|---|-------|
| Referencias..... | 37 |
| Glosario | 39 |
| Anexos..... | - 1 - |
| A Preguntas seleccionadas | - 1 - |
| B Resumen Clústeres (Preguntas y Respuestas)..... | - 5 - |

INDICE DE GRÁFICOS

| | |
|---|----|
| GRÁFICO 1 - VALORACIÓN ATENCIÓN PRIMARIA Y ESPECIALIZADA | 11 |
| GRÁFICO 2 - ¿TUVO LA OPORTUNIDAD DE HACER PREGUNTAS? | 11 |
| GRÁFICO 3 - EXPERIENCIA DE LA ATENCIÓN PRIMARIA..... | 11 |
| GRÁFICO 4 - PREFERENCIA DE ATENCIÓN PRIMARIA Y ESPECIALIZADA..... | 12 |
| GRÁFICO 5 - PREFERENCIA DE LA GESTIÓN DE LA SANIDAD PÚBLICA..... | 12 |
| GRÁFICO 6 - CALIDAD DE CLÚSTER K-MEANS | 21 |
| GRÁFICO 7 - TAMAÑO DE CLÚSTER K-MEANS..... | 21 |
| GRÁFICO 8 - CALIDAD DE CLÚSTER KOHONEN | 23 |
| GRÁFICO 9 - TAMAÑO CLÚSTER KOHONEN..... | 23 |
| GRÁFICO 10 - CALIDAD DE CLÚSTER BIETÁPICO | 25 |
| GRÁFICO 11 - TAMAÑO DE CLÚSTER BIETÁPICO..... | 26 |
| GRÁFICO 12 - IMPORTANCIA DEL PREDICTOR | 27 |
| GRÁFICO 13 - PREGUNTA 1 | 27 |
| GRÁFICO 14 - PREGUNTA 26 | 28 |
| GRÁFICO 15 - MEDIA POR PREGUNTAS..... | 31 |

INDICE DE ILUSTRACIONES

| | |
|---|----|
| ILUSTRACIÓN 1 - CICLO CRISP-DM | 5 |
| ILUSTRACIÓN 2 - MODELO DE DATOS | 15 |
| ILUSTRACIÓN 3 - NODOS DE ORIGEN | 18 |
| ILUSTRACIÓN 4 - NODOS DE OPERACIONES CON INSTANCIAS | 18 |

| | |
|--|----|
| ILUSTRACIÓN 5 - NODOS DE OPERACIONES CON ATRIBUTOS | 18 |
| ILUSTRACIÓN 6 - NODOS DE GRÁFICOS | 18 |
| ILUSTRACIÓN 7 - NODOS DE MODELADOS DE SEGMENTACIÓN | 18 |
| ILUSTRACIÓN 8 - NODOS DE RESULTADOS | 19 |
| ILUSTRACIÓN 9 - NODOS DE EXPORTACIÓN DE DATOS | 19 |
| ILUSTRACIÓN 10 - MODELO | 19 |
| ILUSTRACIÓN 11 - NODOS PRINCIPALES..... | 19 |
| ILUSTRACIÓN 12 - MODELO K-MEANS | 20 |
| ILUSTRACIÓN 13 - MODELO KOHONEN..... | 22 |
| ILUSTRACIÓN 14 - MODELO BIETÁPICO | 24 |
| ILUSTRACIÓN 15 - DISTRIBUCIÓN DE CASILLAS (PREGUNTA 13)..... | 30 |
| ILUSTRACIÓN 16 - DISTRIBUCIÓN DE CASILLAS (PREGUNTAS 26-34, CLÚSTERES 1 Y 3) | 31 |

INDICE DE TABLAS

| | |
|--|----|
| TABLA 1 - PLAN DE PROYECTO..... | 9 |
| TABLA 2 - NOTA ATENCIÓN PRIMARIA | 12 |
| TABLA 3 - PROBLEMAS DE DATOS | 14 |
| TABLA 4 - RESUMEN MODELO K-MEANS..... | 20 |
| TABLA 5 - COCIENTE DE TAMAÑOS DE CLÚSTER K-MEANS | 21 |
| TABLA 6 - ERRORES DE CLASIFICACIÓN K-MEANS..... | 22 |
| TABLA 7 - RESUMEN MODELO KOHONEN | 23 |
| TABLA 8 - COCIENTE DE TAMAÑOS DE CLÚSTER KOHONEN | 24 |
| TABLA 9 - ANÁLISIS KOHONEN..... | 24 |
| TABLA 10 - RESUMEN MODELO BIETÁPICO | 25 |
| TABLA 11 - COCIENTE DE TAMAÑO DE CLÚSTER BIETÁPICO | 26 |
| TABLA 12 - PROPORCIÓN DE LOS MODELOS | 26 |

| | |
|---|----|
| TABLA 13 - RESPUESTAS PREGUNTA 13 | 29 |
| TABLA 14 - PREGUNTADAS DESTACADAS | 30 |

1 Introducción

Motivación

Esta memoria presenta el desarrollo de un proyecto de análisis de datos sanitarios aplicando la metodología CRIPS-DM. Actualmente se realiza anualmente una encuesta a nivel nacional, a los usuarios de la sanidad pública española con el objetivo de realizar un análisis estadístico de la opinión de los mismos con respecto a los servicios prestados por la sanidad pública.

Con el presente proyecto se pretende ampliar el conocimiento obtenido mediante el análisis estadístico, aportando información más detallada de la relación de distintas respuestas de opinión de los usuarios con su calificación final de la sanidad pública, por ejemplo, analizar la importancia, a la hora de valorar los servicios de atención primaria, de que el usuario pueda participar en las decisiones de su tratamiento o que el médico le dé la oportunidad de hacer preguntas o plantear preocupaciones durante las consultas.

También se pretende encontrar distintos patrones de respuestas, que corresponderán a grupos de usuarios de la sanidad pública que opinan de manera similar con respecto a los servicios prestados, y analizar los puntos, sectores y factores que los conducen a opinar positiva o negativamente.

Objetivos

El análisis de datos, desde la perspectiva del desarrollo del software se beneficia de la aplicación de una metodología. El objetivo de este trabajo es el análisis de información de barómetros sanitarios extraídos de la página del Ministerio de Sanidad, Servicios Sociales e Igualdad, siguiendo las directrices de la metodología CRISP-DM, considerada como una de las más recomendables en proyectos de minería de datos, para evaluar el nivel de satisfacción de los usuarios con respecto a los servicios primarios ofrecidos por la sanidad pública española. De manera específica, los objetivos del Trabajo Fin de Grado serían los siguientes:

- Aplicar técnicas de aprendizaje automático para el análisis de información de barómetros sanitarios a partir de datos públicos pertenecientes al Ministerio de Sanidad, Servicios Sociales e Igualdad. Concretamente, se estudiarán y aplicarán técnicas que permitan analizar la segmentación de los grupos de usuarios
- Aplicar una metodología específica de desarrollo de proyectos de minería de datos, con el objetivo de comprobar si el seguimiento de una metodología facilita y ayuda a desarrollar un proyecto de estas características con mayores garantías de calidad y eficiencia.

Organización de la memoria

La memoria consta de los siguientes capítulos:

- **Introducción:** Se describirán las razones que han motivado la realización del proyecto y sus principales objetivos.
- **Estado del arte:** En este capítulo se introducirán los conceptos claves para el desarrollo del proyecto, en este caso la minería de datos y la metodología CRISP-DM.
- **Análisis de datos sanitarios:** En el tercer capítulo, seguiremos todas las fases descritas en la guía de la metodología CRISP-DM, para la realización del proyecto de minería de datos. El capítulo se dividirá según las siguientes fases:
 - Fase de comprensión del negocio.
 - Fase de comprensión de datos.
 - Fase de preparación de datos.
 - Fase de modelado.
 - Fase de evaluación.
 - Fase de implementación.
- **Conclusión y trabajo futuro:** Por último, se concluirá el proyecto con una reflexión sobre el uso de la metodología CRISP-DM y sobre el análisis basado en aprendizaje automático para resolver proyectos de minería de datos, junto con una recomendación para futuros proyectos.

2 Estado del arte

Introducción

En este apartado se describe en qué consiste la minería de datos, así como sus objetivos y los distintos procesos y técnicas más comúnmente utilizados. También se introducirá la metodología CRISP-DM, identificando y explicando sus distintas fases y niveles de abstracción desde un punto de vista teórico, describiendo las distintas tareas a realizar propuestas por la guía de la metodología, las cuales se llevarán a cabo posteriormente en el desarrollo del proyecto.

Minería de datos

La minería de datos es un campo de la estadística y de la informática que se encarga de procesar, analizar y descubrir patrones e información de manera automática o semi-automática en grandes volúmenes de datos. Esos patrones pueden ser vistos como un resumen de los datos estudiados, permitiéndonos dar respuestas a cuestiones anteriormente planteadas, y ser utilizados para análisis descriptivos, predictivos y/o prescriptivos [13] [10]

Los siguientes pasos son procesos del análisis, descubrimiento de la información o actividades en las que se encaja la minería de datos [13] :

- Selección del conjunto de datos.
- Análisis de las propiedades de los datos.
- Transformación de los datos.
- Aplicación de la técnica de minería de datos.
- Extracción del conocimiento.
- Evaluación de los datos obtenidos.

2.1.1 Técnicas de minería de datos

Dentro de las tareas y/o técnicas más representativas de minería de datos podemos encontrar las siguientes [13] :

- Redes Neuronales.
- Árboles de decisión.
- Clustering.
- Regresión Lineal.
- Reglas de asociación (A priori, Carma, Secuencia)

Metodología CRISP-DM

La metodología CRISP-DM (Cross Industry Standard Process for Data Mining) es una de las guías de referencia más utilizadas en proyectos de minería de datos. Incluye descripciones de las fases de un proyecto y las tareas a realizar para cada una de estas, por otro lado, como modelo de proceso ofrece un resumen del ciclo vital de minería de datos. A continuación, se describe cada una de las fases en las que se divide el ciclo vital de la metodología CRISP-DM [8] :

- **Fase de comprensión del negocio**

La primera fase de la metodología se denomina comprensión del negocio o problema, es la fase con mayor importancia ya que en ella se especifican los requisitos y se define el plan a seguir durante las demás fases del proyecto. En esta fase se trata de comprender el problema desde un punto de vista empresarial o institucional para seleccionar la información y los datos que permitan resolverlo. Para la realización de esta fase, CRISP-DM propone seguir las siguientes tareas: determinar el objetivo del negocio, valorar la situación, determinar los objetivos de la minería de datos y por último realizar un plan de proyecto.

- **Fase de comprensión de datos**

La segunda fase se conoce como la comprensión de los datos. En esta fase se recolectan los primeros datos a analizar, con el fin de tener un primer contacto real con el problema planteado en la primera fase y familiarizarse con la información obtenida. En primer lugar, se debe identificar la calidad de los datos y definir una primera hipótesis en base a éstos. Las principales tareas propuestas por la guía en esta fase son: la recolección de datos iniciales, la descripción de los datos, la exploración de los datos y la verificación de la calidad de los datos.

- **Fase de preparación de los datos**

Con los datos ya recopilados, se procede a su preparación para ajustarlos a la técnica de minería de datos que se desea utilizar y construir un conjunto de datos que se ajuste al problema. Esta fase incluye las tareas de selección de datos a los que se aplicará una técnica de modelado, y la limpieza y generación de datos y variables adicionales. Las tareas a realizar en esta fase son: la selección de datos, limpieza de datos, la estructuración de los datos, la integración de los datos y el formateo de los datos.

- **Fase de modelado**

Es el proyecto de minería de datos. Debido a que hay varias técnicas que requieren un formato específico de datos, en muchos casos es necesario volver a la fase de preparación de datos. Las técnicas a utilizar en esta fase se eligen basándose, entre otros, en los siguientes criterios:

- Ser apropiado al problema.
- Cumplir con los requisitos del problema.
- Disponer de datos adecuados.

Las tareas propuestas por la guía son las siguientes: selección de la técnica de modelado, generación de un plan de prueba, construcción del modelo, y evaluación del modelo.

- **Fase de evaluación**

En esta fase se evalúa el modelo seleccionado, y se verifica si se ha cumplido con los criterios de éxito establecidos como objetivo del estudio. Se revisa el proceso y los resultados obtenidos, por si fuera necesario repetir algún paso previo para corregir algún error cometido anteriormente. Si finalmente se concluye que el modelo seleccionado es válido para el estudio, y cumple con los criterios de éxito, se procede a la explotación del modelo. Las tareas que componen esta fase son las

siguientes: evaluación de los resultados, revisión del proceso y determinar próximos pasos.

- **Fase de distribución**

Una vez que el modelo ha sido seleccionado, construido y validado, en esta fase transformaremos el conocimiento y resultados obtenidos en acciones para el negocio. Identificaremos las acciones a realizar por la empresa o cliente, basándonos en los resultados obtenidos a partir del modelo. En esta fase también se documentarán los resultados obtenidos de manera comprensible para el usuario. Las tareas que se llevan a cabo en esta fase son: plan de implantación, plan de monitoreo y mantenimiento, realización del informe final y revisión del proyecto.

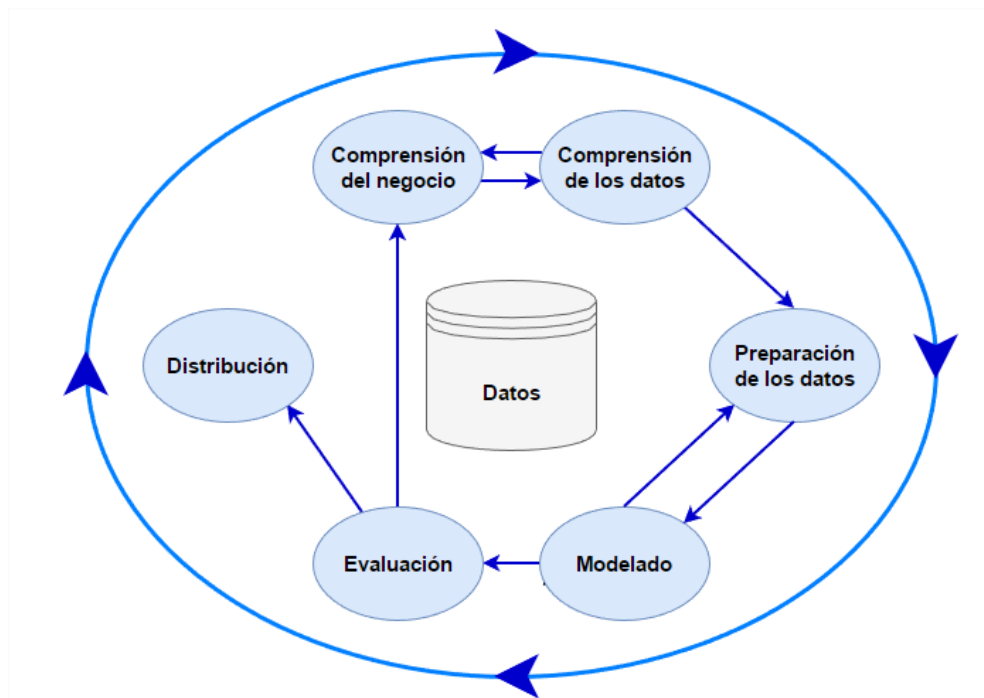


Ilustración 1 - Ciclo CRISP-DM

3 Análisis de Datos Sanitarios

Fase de comprensión del negocio

La primera fase de la metodología CRISP-DM trata de conocer en detalle las razones y objetivos comerciales, para convertirlos en una definición de un problema de minería de datos, y diseñar un plan para alcanzar los objetivos del negocio [8] .

A continuación, se presentan las tareas de la fase de comprensión del negocio, cuyo objetivo es determinar los objetivos y criterios de éxito del proyecto desde una perspectiva empresarial.

3.1.1 Determinar el objetivo del negocio

El objetivo de la investigación y aplicación de la metodología CRISP-DM es evaluar el nivel de satisfacción de los usuarios residentes en España con respecto a la sanidad pública española, principalmente en relación a los centros de atención primaria.

- **Contexto**

Referente a la situación actual del negocio, se dispone de un barómetro sanitario a partir del cual se puede obtener información sobre la percepción y opinión que tienen los usuarios del funcionamiento de la sanidad pública de manera individual.

- **Objetivos del negocio**

El objetivo del negocio es conocer la satisfacción y opinión global de los usuarios para determinar los aspectos positivos y negativos de los servicios sanitarios, y así plantear posibles mejoras aplicables a los mismos.

- **Criterios de éxito del negocio**

Desde una perspectiva comercial, se establece como criterio de éxito del proceso de minería de datos la obtención de información relevante a partir del barómetro sanitario, que nos permita identificar las tareas a realizar para mejorar la experiencia de los usuarios con respecto a los centros de atención primaria.

3.1.2 Valorar la situación

Para la realización del proyecto de minería de datos, disponemos de un barómetro sanitario de ámbito nacional, realizado por el Ministerio de Sanidad, Servicios Sociales e Igualdad, que desde 1993 se realiza cada año mediante un convenio de colaboración con el Centro de Investigaciones Sociológicas (CIS) [3]

- **Inventario de recursos**

En cuanto a recursos software, se dispone de las siguientes herramientas:

- **SPSS Modeler** Premium Academic and Faculty/Author 18.0, una plataforma de análisis predictiva diseñada para realizar predicciones sobre decisiones llevadas a cabo por personas, sistemas o empresas.

- **pgAdmin 4** (versión 1.4). Copyright 2013 - 2017, The pgAdmin Development Team. Herramienta de administración de bases de datos PostgreSQL.
- **PostgreSQL** (64 bits) v9.6.3-1. Sistema de gestión de bases de datos.

Los recursos de hardware de los que disponemos son un ordenador de sobremesa con las siguientes características:

- Procesador: Intel® Core™ i5-3570K CPU @ 3.40GHz
- Memoria instalada (RAM): 8,00 GB
- Tipo de sistema: Sistema operativo de 64 bits, procesador x64
- Sistema Operativo: Windows 10 Home © 2016 Microsoft Corporation
- Capacidad de almacenamiento: 2TB
- Tarjeta gráfica: NVIDIA® GeForce GTX 660

Como fuente de datos se dispone de un fichero en formato texto, que contiene las respuestas dadas por los usuarios, las cuales se encuentran codificadas mediante números enteros.

- **Requisitos, supuestos y restricciones**

Existen restricciones de seguridad relacionadas con el anonimato de las entrevistas realizadas, por lo que no disponemos de los datos personales de los participantes de la encuesta.

- **Riesgos**

Se han identificado una serie de riesgos que podrían surgir en el desarrollo del proyecto. Entre estos riesgos se incluyen los siguientes:

- *Riesgo de plan de proyecto* - La programación de las distintas tareas a realizar en las fases del proyecto podrían durar más de lo planificado. Debido a que se dispone de una licencia temporal de 12 meses de acceso al software (SPSS Modeler) utilizado en el proyecto, existe un riesgo temporal.
- *Riesgo de datos* - Es posible que los datos obtenidos para el proyecto no sean suficientes para concluir los resultados esperados.

- **Terminología**

Ver Glosario de terminología.

- **Costes y beneficios**

Los datos obtenidos a partir del barómetro sanitario no suponen ningún coste adicional al proyecto, debido a que fueron obtenidos de la página del Ministerio de Sanidad, donde se encuentran disponibles a todo el público.

El objetivo del proyecto es conocer la opinión de los usuarios de la sanidad pública con respecto a la atención primaria, e identificar los factores y aspectos que les influyen a crear esa opinión, por tanto, como beneficio se deduce que una vez identificados dichos factores se podrían solucionar y mejorar los aspectos que les lleven a tener una mala consideración de la atención primaria y, por consiguiente, aumentar la satisfacción de los usuarios.

3.1.3 Determinar los objetivos de la Minería de Datos

El objetivo principal de la minería de datos en este proyecto es extraer y seleccionar información del barómetro sanitario, para realizar un estudio y análisis de datos que nos permita encontrar patrones entre las respuestas procedentes de los usuarios de la sanidad pública española. Los datos del barómetro ya incluyen cierto tratamiento estadístico de las respuestas, pero en este trabajo se pretende ampliar el conocimiento identificando aspectos que no se han descubierto en las respuestas.

- **Criterios de éxito de minería de datos**

Los objetivos específicos de la minería de datos son:

- Evaluar el grado de satisfacción de los usuarios de la atención primaria de la sanidad pública española.
- Identificar puntos positivos y negativos del servicio sanitario español, con el objetivo de proponer una serie de acciones a realizar que permitan mejorar la experiencia de los usuarios.
- Determinar mediante la tarea de segmentación los grupos de usuarios que presentan una opinión similar respecto a la atención primaria, tratando de descubrir características comunes a todo el grupo que permitan analizar si alguna o todas esas características son determinantes a la hora de conducir a un usuario a expresar la opinión que expresa.

3.1.4 Realizar un plan de proyecto

El coste de personal del proyecto será de una persona, y se dividirá en las siguientes fases estimadas:

| Fase | Tiempo | Riesgo |
|--------------------------|-----------|--|
| Comprensión del negocio | 2 semanas | Ninguno |
| Comprensión de los datos | 4 semanas | Problema de datos, problemas tecnológicos |
| Preparación de los datos | 5 semanas | Problema de datos, problemas tecnológicos |
| Modelado | 4 semanas | Problemas tecnológicos, dificultad para encontrar el modelo adecuado |
| Evaluación | 2 semanas | Dificultad para evaluar resultados ambiguos |
| Distribución | 1 semana | Dificultad para un informe final de fácil comprensión |

Tabla 1 - Plan de proyecto

Fase de comprensión de datos

En esta fase se seleccionan y se estudian en profundidad los primeros datos recogidos, con el fin de familiarizarnos con ellos y tener un primer contacto real con el problema planteado en la primera fase [8] .

3.1.5 Recolección de datos iniciales

El barómetro sanitario es un estudio de opinión que se realiza de manera anual desde 1993 por el Ministerio de Sanidad, Servicios Sociales e Igualdad junto al Centro de Investigación Sociológica. Está diseñado para conocer el nivel de satisfacción de los usuarios del servicio sanitario público y obtener una opinión sobre medidas relacionadas con la política sanitaria [4]

Los resultados del estudio se obtienen mediante un cuestionario que se realiza de manera personal en el domicilio del encuestado, los cuales han sido seleccionados por rutas aleatorias de ubicación y cuotas de sexo y edad. Cada año, se asignan 250 entrevistas a cada una de las Comunidades Autónomas, el resto (3050 entrevistas) se distribuye proporcionalmente al tamaño de la población residente en cada una de ellas. Este diseño permite conocer la opinión de los usuarios mayores de 18 años sobre los servicios de sanidad pública. Se han realizado 7800 entrevistas por año, distribuidas en tres submuestras de 2600 entrevistas cada una.

3.1.6 Descripción de los datos

El barómetro sanitario es un estudio de opinión que se realiza cada año con la finalidad de conocer la opinión de los ciudadanos para tomar en consideración sus expectativas, y mejorar el sistema nacional de salud, como elemento importante para establecer las prioridades de las políticas de salud.

La población del estudio corresponde a personas de ambos sexos mayores de 18 años, residentes en las 17 comunidades y en las 2 ciudades autónomas. El estudio recoge la opinión de los ciudadanos sobre el funcionamiento del sistema sanitario público español y su grado de satisfacción con respecto a la calidad de las prestaciones de los distintos servicios sanitarios públicos [2]

El cuestionario contiene entre 60 y 70 preguntas, y está dividido en cinco partes principales. En la primera parte se aborda una valoración general del funcionamiento del Sistema Nacional de Salud, en la que se recogen distintos aspectos como la valoración de los servicios sanitarios, las listas de espera, la equidad en el acceso a los servicios, etc. Las otras cuatro partes se centran en servicios específicos de la sanidad: atención primaria, servicios especializados, urgencias y hospitalización.

Las respuestas de las encuestas están recogidas en un fichero de texto, donde cada línea representa un entrevistado. Sus respuestas se encuentran codificadas en números enteros. Disponemos además de un fichero PDF en el que tenemos la traducción del código de cada respuesta codificada.

3.1.7 Exploración de los datos

Al hacer un primer análisis de los datos obtenidos en el barómetro sanitario, podemos observar (Gráfico 1 – izquierda) que alrededor del 86% de los entrevistados valora la asistencia recibida en la atención primaria (médico de familia o pediatra) como buena o muy buena, mientras que alrededor del 14% la valora como regular, mala o muy mala. La relación paciente/médico se valora positivamente, ya que el 88% de los encuestados pudieron preguntar y resolver las dudas sobre su salud o su tratamiento (Gráfico 2 - izquierda). El 23% de los usuarios manifiestan que su experiencia ha sido mejor o mucho mejor de lo esperado, mientras que el 8% opina que fue peor o mucho peor, el 69% restante opina que la atención ha sido igual a como esperaban (Gráfico 3).



Gráfico 1 - Valoración atención primaria y especializada



Gráfico 2 - ¿Tuvo la oportunidad de hacer preguntas?



Gráfico 3 - Experiencia de la atención primaria

Por otro lado, en la atención especializada podemos observar (Gráfico 1 - derecha) que las valoraciones de muy buena o buena bajan a un 80%, mientras que el 16% la valora como regular, mala o muy mala (el porcentaje restante corresponde a respuestas de tipo “no sabe, no contesta”). La relación paciente/médico también se ha valorado positivamente, el 85% de los encuestados pudieron resolver sus dudas y hacer preguntas sobre su salud o su tratamiento (Gráfico 2 - derecha).



Gráfico 4 - Preferencia de Atención Primaria y Especializada

También observamos, de acuerdo al Gráfico 4, que los usuarios entrevistados prefieren los servicios sanitarios de los centros públicos a los de los centros privados. En relación con la atención primaria, un 75% prefiere los servicios públicos frente un 25% que prefiere los servicios privados, pero esta diferencia disminuye cuando hablamos de la atención especializada, donde el 57% sigue prefiriendo acudir a un centro público mientras que el 36% prefiere uno privado.

Por otra parte, la mayoría de ciudadanos (69%) considera que la gestión de la sanidad pública debe ser llevada a cabo por la propia Administración Pública mientras que el 13% opina que las empresas privadas pueden realizar una mejor gestión (Gráfico 5).

Gestión de la Sanidad Pública

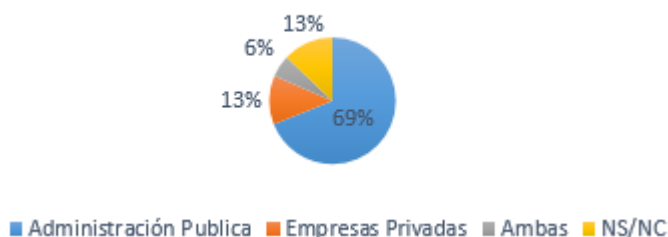


Gráfico 5 - Preferencia de la Gestión de la Sanidad Pública

| Nota a la Atención Primaria | Cantidad Total | Porcentaje | No acudieron en los último 12 meses | Porcentaje | Porcentaje Válido (Cantidad Total - No acudieron en los últimos 12 meses) | Grupos |
|-----------------------------|----------------|------------|-------------------------------------|------------|---|--------|
| 10 | 777 | 5% | 178 | 1% | 4% | 40% |
| 9 | 1237 | 8% | 288 | 2% | 6% | |
| 8 | 3191 | 21% | 926 | 6% | 15% | |
| 7 | 3364 | 22% | 986 | 6% | 15% | |
| 6 | 2564 | 17% | 736 | 5% | 12% | |
| 5 | 2119 | 14% | 602 | 4% | 10% | 32% |
| 4 | 1004 | 7% | 268 | 2% | 5% | |
| 3 | 519 | 3% | 141 | 1% | 2% | |
| 2 | 229 | 1% | 71 | 0% | 1% | |
| 1 | 370 | 2% | 78 | 1% | 2% | |
| Total | 15374 | 100% | 4274 | 28% | 72% | 72% |

Tabla 2 - Nota Atención Primaria

Como se puede observar en la Tabla 2, de los usuarios que han valorado numéricamente los servicios prestado en la asistencia primaria, un 28% de ellos no han acudido a los centros en los últimos 12 meses. De los demás usuarios que sí han acudido a los centros de atención primaria en los últimos 12 meses, podemos observar que el 40% de ellos han valorado los servicios prestados con una nota superior a 7, mientras que el 32% los han valorado con un 6 o menos.

3.1.8 Verificación de la calidad de los datos

Debido a que los datos fueron obtenidos del barómetro sanitario realizado por el Ministerio de Sanidad, Servicios Sociales e Igualdad, mediante un convenio de colaboración con el Centro de Investigaciones Sociológicas, no se espera un alto número de errores en los resultados de las encuestas. Todos los resultados de las encuestas se encuentran codificados en números enteros, para facilitar su manejo y análisis, disminuyendo la probabilidad de errores a la hora de ser procesados.

Entre las respuestas obtenidas en las encuestas del estudio, nos encontramos con un elevado número de respuestas “no sabe” y “no contesta”, que podrían afectar al resultado del estudio. Esto se debe a que el barómetro sanitario es un estudio de opinión, en el que el usuario tiene la posibilidad de no contestar a determinadas preguntas. En caso de que esto influya en los resultados, la solución propuesta es asignar el valor de la media de las demás respuestas para la pregunta en cuestión, o mantener estos valores, de NS/NC, para que generen un sector de la población de datos.

Fase de preparación de los datos

Una vez finalizada la fase de recolección y exploración, pasamos a la fase de preparación de los datos, donde los adaptaremos y ajustaremos a la técnica de minería de datos que se utilizará en el estudio. En esta fase seleccionaremos y limpiaremos los datos para la siguiente fase de modelado. [8]

Para la preparación de los datos, se han introducido en una base de datos SQL. El gestor elegido es PostgreSQL por ser de libre distribución e incorporar las características necesarias para realizar este proyecto. La herramienta de acceso y manipulación de la base de datos utilizada es la versión 1.4 de pgAdmin.

3.1.9 Selección de datos

En la encuesta, se dispone de una serie de preguntas que abarcan varios sectores de la sanidad pública española, entre ellos encontramos los servicios primarios, servicios especializados, urgencias y hospitalización. En esta fase, seleccionaremos las preguntas generales, y las relacionadas únicamente con los servicios primarios de la sanidad, debido a que el alcance del proyecto no incluye los demás sectores. De esta forma el cuestionario de preguntas de interés para el estudio de este trabajo se reduce a 44 preguntas. A partir de las preguntas seleccionadas, se ha introducido en la base de datos todos los datos de las encuestas realizadas en los años 2015 y 2016, un total de 15498 encuestas.

- Ver Anexo A, con el listado de preguntas.

3.1.10 Limpieza de datos

Como se ha descrito en la actividad de Verificación de la calidad de los datos, debido a que el barómetro fue realizado por el Ministerio de Sanidad junto al Centro de Investigaciones Sociológicas, no se han encontrado un elevado número de problemas en los datos seleccionados. Tras la limpieza de datos, el número de encuestas disponibles no se ha reducido. Algunos de los problemas encontrados son los siguientes:

| Tipo de problema | Descripción | Solución |
|-------------------------------|--|--|
| Datos perdidos | Encuestas sin código de municipio. | Se ha generado en la base de datos por cada provincia un código de municipio 0 que equivale a "N/A". Esto se utilizará cuando no venga informado el código del municipio. |
| Metadatos ausentes o erróneos | Código municipio erróneo. El código de municipio no coincide con ninguno de los municipios listado en el barómetro | Se ha analizado cada caso por separado y se le ha asignado a la encuesta el código de un municipio existente (reemplazando el erróneo). |
| Incoherencias de codificación | Algunas preguntas dependían de la respuesta dada en una pregunta anterior, por lo que en algunos casos venían con código 0 (no corresponde a ninguna posible respuesta). | A estas respuestas, como no aplican debido a una contestación anterior, se le ha asignado valor nulo (null). Esta solución no anula la encuesta, pero la respuesta no afecta en el estudio, a nivel estadístico. |

Tabla 3 - Problemas de datos

3.1.11 Estructuración de los datos

Para la estructuración de los datos se ha generado una base de datos siguiendo un esquema de tipo copo de nieve. El modelo copo de nieve es similar al modelo en estrella, contiene una única tabla central, a la que llamamos “tabla de hechos”, y varias tablas de N dimensiones alrededor. Las tablas de dimensiones representan los factores por lo que se analiza una determinada área del negocio, contienen los metadatos de la tabla de hechos. Por otro lado, la tabla de hechos representa el objeto del análisis, contiene los datos del estudio [14].

Lo que diferencia el modelo estrella del modelo copo de nieve son las dimensiones ya que, en el modelo en estrella se implementa cada dimensión con una única tabla, mientras que en el modelo copo de nieve se pueden implementar con más de una tabla de datos.

La finalidad de este modelo es des-normalizar las tablas. Por otro lado, este esquema es ideal por su simplicidad y velocidad cuando se utiliza para análisis multidimensionales.

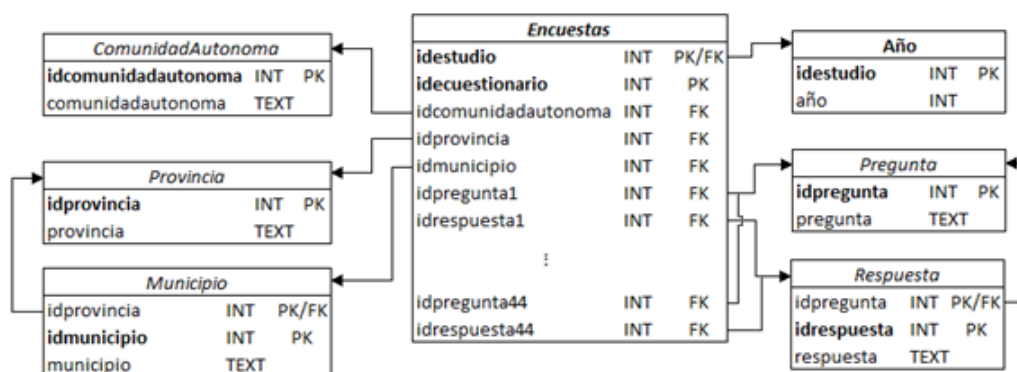


Ilustración 2 - Modelo de datos

A partir de los ficheros Datos_2016 y Datos_2015, que contienen los resultados del barómetro sanitario en formato texto, tal y como pueden descargarse de la página web del Ministerio de Sanidad, con cada posición codificada con un número entero, se ha generado una tabla de hechos *Encuestas* y 6 tablas dimensión a su alrededor.

- **Encuestas**: Se trata de la tabla de hechos del modelo de datos. En ella se guarda, por cada línea, toda la información necesaria para el estudio. La información está indicada con una clave foránea que apunta a su respectiva tabla de dimensiones, la cual define las características del valor contestado por el entrevistado, por ejemplo, la comunidad autónoma, la provincia, o la respuesta a una pregunta determinada. La tabla de hechos registra eventos específicos, por lo que en este caso registra una encuesta concreta realizada a un entrevistado.
- **Comunidad Autónoma**: Tabla de dimensión geográfica. Define datos de las Comunidades Autónomas a las que pertenecen los entrevistados.
- **Provincia**: Tabla de dimensión geográfica. Define datos de las provincias a las que pertenecen los entrevistados.
- **Municipio**: Tabla de dimensión geográfica. Define datos de los municipios a los que pertenecen los entrevistados.
- **Año**: Tabla de tiempo. Define el año en el que se ha realizado la encuesta.
- **Pregunta**: Tabla de preguntas. Define las preguntas de la encuesta seleccionadas para el estudio.
- **Respuesta**: Tabla de respuestas. Define todas las posibles respuestas a cada pregunta realizada en la encuesta.

3.1.12 Integración de los datos

Dentro de la encuesta realizada por el Ministerio de Sanidad no existe ninguna referencia temporal relacionada a cuándo se ha realizado, aunque desde su página web se pueden descargar los microdatos de las encuestas separados por años. Para este estudio, hemos descargado los datos de las encuestas realizadas durante los años 2015 y 2016, para

disponer de un mayor número de respuestas y poder ajustar con más detalle la información. El dato del año en el que se ha realizado la encuesta se ha integrado en la base de datos mediante la tabla “Año”, en el que relacionamos el número del estudio a un año en concreto (2015 o 2016) [3]. Actualmente esta tabla contiene únicamente dos filas, pero en trabajos futuros se pretende añadir más años, e incluso dividirla en semestre, cuatrimestre o meses.

Para la carga de los datos en la base de datos se ha realizado un pequeño programa Java, que lee los ficheros Datos_2015 y Datos_2016, y por cada línea leída (que representa una encuesta completa a un usuario) se genera una sentencia SQL “INSERT”, y la almacena en un script, que posteriormente se ejecutará en el gestor de bases de datos.

3.1.13 Formateo de los datos.

La fuente de datos de la que disponemos para el estudio es el barómetro sanitario, que contiene respuestas de los entrevistados en formato texto. Las respuestas están representadas mediante una codificación numérica. Debido a eso, ha sido necesario formatear los datos de las encuestas a sus respectivos números enteros para la inserción a la base de datos. Esta actividad también facilitará el modelado de los datos.

Fase de modelado

En esta fase de la metodología CRISP-DM seleccionaremos algunas de las técnicas de modelado disponibles en la herramienta SPSS Modeler, que sean más apropiadas para los objetivos definidos en el alcance del proyecto. A continuación, realizaremos un plan de pruebas de ejecución de los modelos seleccionados, para proceder después a aplicar dichas técnicas [8].

3.1.14 Selección de la técnica de modelado

Para el modelado disponemos de la herramienta SPSS Modeler, que tiene incorporadas una serie de técnicas separadas por las categorías: Analytic Server, Clasificación, Asociación y Segmentación. Debido a que el objetivo del negocio es conocer el grado de satisfacción de los usuarios de la atención primaria de la sanidad pública española, se ha decidido utilizar una tarea de segmentación. La razón es que en la segmentación se trata de descubrir clústeres o segmentos en los datos de los que se dispone, siendo una tarea descriptiva, mientras que la clasificación trata de predecir la clase a la que pertenecen unos datos nuevos. Como el barómetro contiene todos los datos, el objetivo es descubrir los clústeres para ver si revelan datos esclarecedores sobre la opinión de los usuarios respecto a la atención primaria, como ya se ha comentado en los objetivos del negocio.

Las técnicas de modelado basadas en segmentación son técnicas que aplican algoritmos de agrupamiento. Estas técnicas siguen un procedimiento de agrupación de vectores (clúster) de acuerdo con un criterio específico, que suele ser la distancia entre los valores de los datos [12]. Las técnicas de segmentación seleccionadas para este proyecto son:

- **K-Means:** Algoritmo que inicialmente agrupa un conjunto de patrones en un número dado de subconjuntos llamados “clúster” y, por cada iteración, reasigna

algunos de los patrones de un clúster a otros clústeres, hasta que se cumple una condición de finalización [7] .

- **Kohonen:** Los mapas de Kohonen son mapas auto-organizados que, son entrenados usando un algoritmo de aprendizaje no supervisado y generan una representación discreta del espacio de las muestras de entrada [7] .
- **Bietápico:** Se utiliza para agrupar el conjunto de datos en grupos distintos cuando se desconoce su significado, intentando revelar patrones en el conjunto de campos de entrada [7] .

3.1.15 Generación de un plan de prueba

Como paso final antes de generar los modelos, crearemos un procedimiento para la comprobación de la calidad y validez de cada técnica. Para dicha comprobación, tendremos en cuenta el resumen del modelo generado por la herramienta SPSS Modeler, donde se especifica una instantánea del modelo, incluyendo una medida de silueta de la cohesión y distancia entre clústeres, indicando si los resultados son pobres, correctos o buenos.

Este resumen nos permite comprobar de manera rápida si la calidad del modelo es insuficiente. Un resultado “bueno” indica que los datos resultantes de la aplicación de la técnica reflejan una evidencia razonable o sólida de que existe una estructura de clústeres, un resultado “correcto” indica que la evidencia es débil, y en el último caso cuando el resultado es “pobre” indica que no hay evidencias obvias [7]

La medida de silueta se calcula a partir de todas las instancias de origen (los datos), aplicando la siguiente fórmula:

$$\frac{(B - A)}{\max(A, B)}$$

Donde A es la distancia de la instancia al centro del clúster, y B es la distancia de la instancia al centro del clúster más cercano al que no pertenece. Por lo que un valor de silueta igual a 1 significa que la instancia se encuentra ubicada en el centro de su clúster, mientras que si la silueta vale -1, la instancia se encuentra en el centro del clúster más cercano al que no pertenece. Un valor 0 significa que la instancia se encuentra a equidistancia del centro de su propio clúster y del siguiente clúster más cercano [7]

Como plan de prueba, se van a realizar pruebas de caja negra. Se realizarán pruebas funcionales de la ejecución de la técnica de modelado, observando los datos de entrada y los resultados obtenidos, comprobando que estos resultados son consecuentes con los datos estadísticos obtenidos en la exploración de los datos [15] .

3.1.16 Construcción del modelo

En este punto, tenemos preparados los datos y la información necesaria para generar los modelos que hemos considerado en la actividad de selección de la técnica de modelado. Para la construcción del modelo disponemos de distintos “nodos” que utilizaremos en el workspace de la herramienta SPSS Modeler, que sirven para diversas funciones, por ejemplo:

- **Nodos de origen:** nodos que permiten traer información externa a la herramienta, a partir de distintas fuentes, por ejemplo, desde una base de datos, de un archivo de texto, de una hoja de Excel, XML, etc.



Ilustración 3 - Nodos de origen

- **Nodos de operaciones con instancias:** nodos que permiten realizar distintas operaciones con las instancias, por ejemplo, seleccionar, ordenar, equilibrar, agregar, etc.

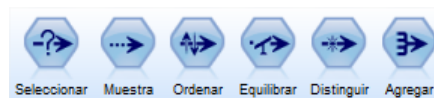


Ilustración 4 - Nodos de operaciones con instancias

- **Nodos de operaciones con atributos:** nodos que sirven para realizar operaciones con atributos específicos de los datos, por ejemplo, filtrar, derivar, rellenar, anonimizar, etc.



Ilustración 5 - Nodos de operaciones con atributos

- **Nodos de gráficos:** nodos que generan gráficos a partir de datos de otros nodos.

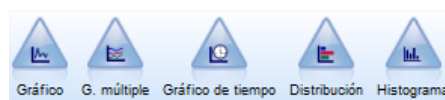


Ilustración 6 - Nodos de gráficos

- **Nodos de modelado:** nodos que aplican técnicas de modelado a los datos seleccionados en otros nodos. Se dividen en cuatro categorías: Analytic Server, Clasificación, Asociación y Segmentación.

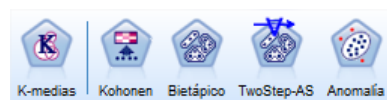


Ilustración 7 - Nodos de modelados de segmentación

- **Nodos de resultados:** nodos con distintos métodos para mostrar resultados obtenidos con la aplicación de técnicas de modelado, por ejemplo, tablas y matrices.

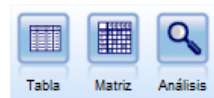


Ilustración 8 - Nodos de resultados

- **Nodos de exportación de datos:** nodos que nos permiten exportar los datos fuera de la herramienta en distintos formatos.



Ilustración 9 - Nodos de exportación de datos

- **Modelos:** Modelo generado a partir de los datos y la técnica de modelado seleccionada.



Ilustración 10 - Modelo

Para la construcción de los tres modelos seleccionados, haremos uso de dos nodos para importar y leer los datos de la base de datos preparada en la fase anterior de la metodología. Utilizaremos un nodo de “origen” base de datos SQL, que actuará como el dataset (tabla de hechos) del método de segmentación, y lo conectaremos a un nodo “tipo” (nodo de operación con campos) para leer e identificar los valores de los datos introducidos. El nodo “tipo” se encargará de leer los datos del dataset para identificar si son valores continuos, nominales, categóricos, etc., para tenerlo en cuenta cuando se realicen cálculos de distancia. También podremos seleccionar las columnas del dataset a utilizar para la técnica de modelado (filtrado de columnas). Esto se hará para excluir las columnas que identifican las preguntas (idpreguntaX) y las que identifican la encuesta (idEstudio e idCuestionario), ya que no se deben tener en cuenta para la segmentación. A partir del nodo “tipo”, conectaremos los distintos nodos de técnica de modelado.



Ilustración 11 - Nodos principales

Como en esta fase se ha optado por probar tres técnicas distintas, realizaremos la construcción de los modelos separándolos en tres apartados, uno por cada técnica de modelado.

3.1.16.1 K-Means

Método para análisis de clústeres. Se utiliza para agrupar conjuntos de datos en distintos clústeres que en un principio se desconocen. Este modelo intenta revelar los patrones en el

conjunto de datos, agrupando las instancias de manera que las de un mismo clúster tienden a ser similares entre ellas [7] .

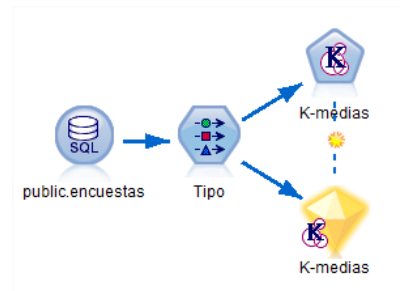


Ilustración 12 - Modelo K-Means

- **Configuración de parámetros**

Para configurar los parámetros de entrada, primero introducimos el nodo del modelo K-Means conectado al nodo “tipo” previamente introducido, posteriormente damos doble clic sobre el nodo del modelo para abrir la ventana de configuración, donde asignaremos los siguientes parámetros de creación:

- Utilizar los datos en particiones: falso
- Calcular puntuaciones de propensión en bruto: falso
- Calcular puntuaciones de propensión ajustada: falso
- Número de clústeres: 3
- Generar campo de distancia: falso
- Etiqueta de clúster: Cadena
- Prefijo de etiqueta: clúster
- Optimizar: Memoria
- Modo: Simple

- **Ejecución y descripción del modelo**

Una vez introducido el nodo del modelo y configurados los parámetros de creación, ejecutamos el nodo dando click el botón “play” de la ventana de configuraciones para producir los resultados visibles del modelo.

Al abrir los resultados, a la izquierda veremos un resumen del modelo junto a un gráfico de calidad de cluster. El resumen nos informa del algoritmo ejecutado, el número de entradas leídas (número de columnas seleccionadas por su importancia en el modelo, referentes a las preguntas del barómetro) y el número de clústeres encontrados:

| Resumen del modelo | |
|--------------------|----------|
| Algoritmo | K-medias |
| Entradas | 18 |
| Clústeres | 3 |

Tabla 4 - Resumen modelo K-Means

El gráfico de calidad de clúster nos indica el resultado del cálculo de la silueta de cohesión y separación del clúster. En él podemos distinguir tres bloques que sirven para calificar la calidad de los clústeres encontrados, donde una silueta de resultado -1 a 0,2 será calificada como “mala”, de 0,2 a 0,5 “regular” y de 0,5 a 1 “buena”. Como podemos observar, el modelo K-Means genera un conjunto de clústeres con calidad regular, con una silueta de valor promedio 0,5.

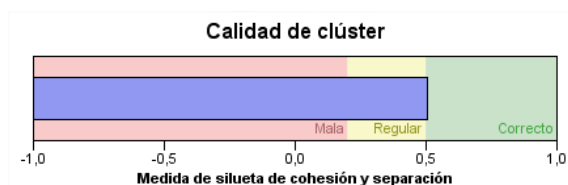


Gráfico 6 - Calidad de clúster K-Means

En la parte de la derecha del resumen generado podemos ver un gráfico que muestra el tamaño proporcional (en porcentaje) de los clústeres identificados por el modelo, y una tabla que nos informa de los tamaños del clúster más pequeño y del clúster más grande, junto a un valor de cociente de tamaño entre ellos, la proporción entre uno y el otro:

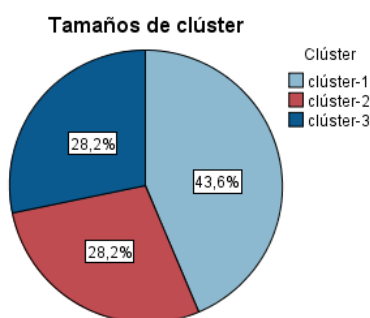


Gráfico 7 - Tamaño de clúster K-Means

| | |
|--|--------------|
| Tamaño del clúster más pequeño | 4367 (28,2%) |
| Tamaño del clúster más grande | 6758 (43,6%) |
| Cociente de tamaños: De clúster más grande a clúster más pequeño | 1,55 |

Tabla 5 - Cociente de tamaños de clúster K-Means

A primera vista, podemos observar que los porcentajes obtenidos con este modelo se asemejan a los porcentajes analizados en la Tabla 2 de la tarea Exploración de los datos.

Un dato más del que disponemos en el modelo K-Means es la Tabla 6, con un cálculo de error de clasificación para cada iteración realizada en el modelo, hasta llegar al criterio de convergencia asignado en los parámetros del propio modelo que, en este caso, debido a que se ha ejecutado en modo “Simple”, son 20 iteraciones. Como podemos observar, el error de clasificación va disminuyendo con cada iteración.

| Número de clústeres: 3 | | | |
|------------------------|-------|-----------|-------|
| Iteración | Error | Iteración | Error |
| 1 | 1,465 | 11 | 0,044 |
| 2 | 0,687 | 12 | 0,043 |
| 3 | 0,558 | 13 | 0,042 |
| 4 | 0,674 | 14 | 0,013 |
| 5 | 0,37 | 15 | 0,008 |
| 6 | 0,199 | 16 | 0,008 |
| 7 | 0,116 | 17 | 0,003 |
| 8 | 0,074 | 18 | 0,002 |
| 9 | 0,069 | 19 | 0,001 |
| 10 | 0,058 | 20 | 0,002 |

Tabla 6 - Errores de clasificación K-Means

3.1.16.2 Kohonen

Los mapas de Kohonen son un tipo de red neuronal que realiza agrupaciones en clústeres de los conjuntos de datos, cuando no se sabe lo que son al principio. Las instancias agrupadas en un mismo clúster tienden a ser similares entre ellas. Las unidades básicas del modelo son neuronas que se organizan en dos capas, una de entrada y otra de salida. Las neuronas de entrada están conectadas a las de salida, y cada conexión tiene una “fuerza” que, durante la fase de entrenamiento, compete con las demás conexiones para ganar cada instancia y ajustar los pesos [7] .

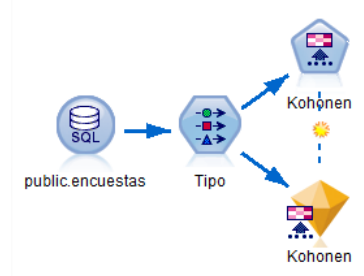


Ilustración 13 - Modelo Kohonen

- **Configuración de parámetros**

Para configurar los parámetros de entrada, primero introducimos el nodo del modelo Kohonen conectado al nodo “tipo” previamente introducido, posteriormente damos doble clic sobre el nodo del modelo para abrir la ventana de configuración, en la que asignaremos los siguientes parámetros de creación:

- Utilizar los datos en particiones: falso
- Calcular puntuaciones de propensión en bruto: falso
- Calcular puntuaciones de propensión ajustada: falso
- Continuar entrenando modelo existente: falso

- Mostrar gráfico de retroalimentación: true
- Detener cuando: Valor predeterminado
- Asignación de partición repetible: falso
- Utilizar codificación de conjuntos binaria: falso
- Optimizar: Memoria
- **Ejecución y descripción del modelo**
Una vez introducido el nodo del modelo y configurado los parámetros de creación, ejecutamos el nodo para producir los resultados visibles, de manera similar a la ejecución del modelo anterior:

| Resumen del modelo | |
|--------------------|---------|
| Algoritmo | Kohonen |
| Entradas | 9 |
| Clústeres | 12 |

Tabla 7 - Resumen modelo Kohonen

De igual forma, el resumen nos muestra una tabla informando del algoritmo utilizado para la generación del modelo, el número de entradas leídas (número de columnas seleccionadas por su importancia en el modelo, referentes a las preguntas del barómetro) y el número de clústeres identificados. El siguiente dato que nos muestra es el gráfico de calidad del clúster que mide el promedio de la silueta de separación entre clústeres. En este modelo podemos observar que la calidad de los clústeres es mala, ya que el promedio de la silueta tiene un valor de 0,2.

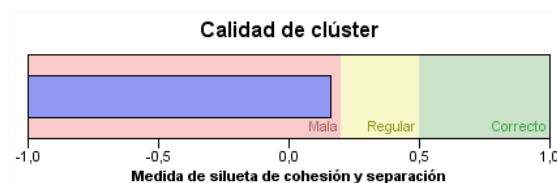


Gráfico 8 - Calidad de clúster Kohonen

A la derecha del resumen generado vemos el gráfico circular con el tamaño proporcional de los 12 clústeres identificados con el modelo, y una tabla con el dato del tamaño del clúster más pequeño, el tamaño del clúster más grande y la proporción que hay entre ellos.

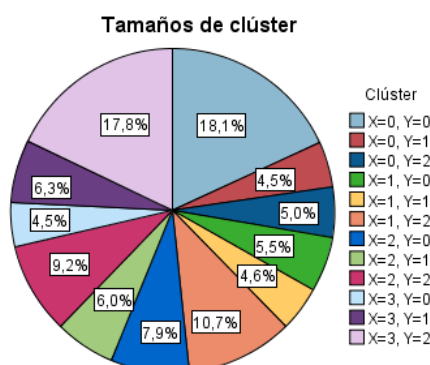


Gráfico 9 - Tamaño clúster Kohonen

| | |
|--|--------------|
| Tamaño del clúster más pequeño | 690 (4,5%) |
| Tamaño del clúster más grande | 2798 (18,1%) |
| Cociente de tamaños: De clúster más grande a clúster más pequeño | 4,06 |

Tabla 8 - Cociente de tamaños de clúster Kohonen

Como dato extra, en el modelo Kohonen podemos ver la siguiente tabla de análisis que nos muestra las dimensiones del mapa de Kohonen (KX y KY) y las dos capas, la de entrada con 17 neuronas y la de salida con 12.

| Análisis | |
|-----------------|-------------|
| \$KX-Kohonen | 4 |
| \$KY-Kohonen | 3 |
| Capa de entrada | 17 neuronas |
| Capa de salida | 12 neuronas |

Tabla 9 - Análisis Kohonen

3.1.16.3 Bietápico

Igual que los modelos Kohonen y K-Means, los modelos de clústeres biatápicos intentan revelar los patrones en el conjunto de campos de entrada. Las instancias se agrupan de manera que las de un mismo clúster tienden a ser similares entre ellas. Este modelo se realiza en dos pasos, en el primer paso comprime los datos de entrada en un conjunto de subclústeres, y en el segundo paso se utiliza un método de agrupación de clústeres jerárquico para introducir los subclústeres en clústeres cada vez más grandes [7] .

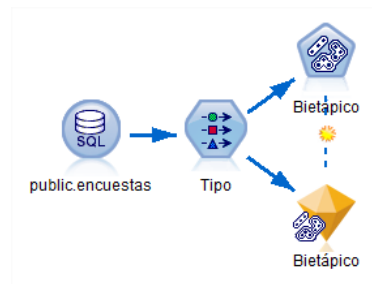


Ilustración 14 - Modelo Bietápico

- **Configuración de parámetros**

Para configurar los parámetros de entrada, primero introducimos el nodo del modelo Bietápico conectado al nodo “tipo” previamente introducido. Posteriormente damos doble clic sobre el nodo del modelo para abrir la ventana de configuración, donde asignaremos los siguientes parámetros de creación:

- Utilizar los datos en particiones: falso
- Calcular puntuaciones de propensión en bruto: falso
- Calcular puntuaciones de propensión ajustada: falso
- Estandarizar campos numéricos: true

- Excluir valores atípicos: falso
- Etiqueta de clúster: Cadena
- Prefijo de etiqueta: clúster
- Calcular automáticamente número de clústeres: true
- Número máximo de clústeres: 15
- Núm. mínimo de clústeres: 3
- Medida de la distancia: Log-verosimilitud
- Criterio de agrupación en clústeres: Criterio bayesiano de Schwarz (BIC)

- **Ejecución y descripción del modelo**

Una vez introducido el nodo del modelo y configurado los parámetros de creación, ejecutamos el nodo para producir los resultados visibles, de manera similar a la ejecución de los modelos anteriores:

| Resumen del modelo | |
|--------------------|-----------|
| Algoritmo | Bietápico |
| Entradas | 23 |
| Clústeres | 3 |

Tabla 10 - Resumen modelo Bietápico

El resumen nos muestra una tabla informando del algoritmo utilizado para la generación del modelo, el número de entradas leídas (número de columnas seleccionadas por su importancia en el modelo, referentes a las preguntas del barómetro) y el número de clústeres identificados. Posteriormente nos muestra el gráfico de calidad del clúster que mide el promedio de la silueta de separación entre clústeres. En este modelo podemos observar que la calidad de los clústeres es regular, ya que el promedio de la silueta tiene un valor de 0,3.

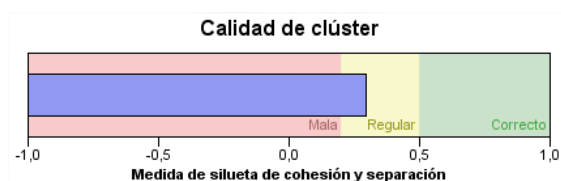


Gráfico 10 - Calidad de clúster Bietápico

A la derecha del resumen generado vemos el gráfico circular con el tamaño proporcional de los 3 clústeres identificados con el modelo, y una tabla con el dato del tamaño del clúster más pequeño, el tamaño del clúster más grande y la proporción que hay entre ellos.

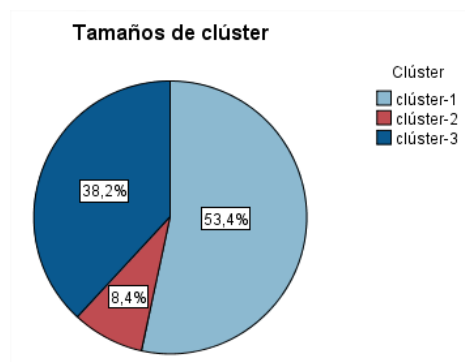


Gráfico 11 - Tamaño de clúster Bietápico

| | |
|--|--------------|
| Tamaño del clúster más pequeño | 665 (8,4%) |
| Tamaño del clúster más grande | 4226 (53,4%) |
| Cociente de tamaños: De clúster más grande a clúster más pequeño | 6,35 |

Tabla 11 - Cociente de tamaño de clúster Bietápico

3.1.17 Evaluación del modelo.

Una vez ejecutadas las tres técnicas de modelado, pasamos a la selección del modelo generado que utilizaremos para encontrar y definir patrones en los usuarios encuestados. Basándonos en el criterio que hemos seleccionado para evaluar la calidad del modelo (recomendado por la herramienta SPSS Modeler), la medida de silueta de cohesión y separación de los clústeres, el número de clústeres generados y su estructura, seleccionaremos para el estudio el modelo de datos K-Means.

Podemos observar que el modelo K-Means ha generado tres clústeres de distintos tamaños, uno que es mayor que los otros, y dos que tienen un tamaño similar. Estos porcentajes se asemejan a los porcentajes analizados en la Tabla 2 de la tarea de Exploración de los datos, donde el clúster más grande podría corresponder a los usuarios que han acudido a los servicios de atención primaria en los últimos 12 meses y los han valorado positivamente (nota 7 o superior), mientras que los dos clústeres de menor tamaño, uno correspondería a los usuarios que han acudido en los último 12 meses y han valorado negativamente (6 o inferior), y el otro clúster serían los usuarios que no han acudido en los último 12 meses, por lo que su nota no tiene una alta relevancia para el estudio.

| Valor | Proporción | % | Recuento |
|-----------|------------|-------|----------|
| clúster-1 | | 43.61 | 6758 |
| clúster-2 | | 28.18 | 4367 |
| clúster-3 | | 28.22 | 4373 |

Tabla 12 - Proporción de los modelos

Con esta información, podemos suponer que existen tres patrones en las respuestas de los entrevistados. Para comprobar que el modelo es válido, y concluir que cada uno de los clústeres corresponde a un conjunto de personas con opiniones similares sobre la sanidad pública y servicios primarios, nos fijaremos en el grado de importancia asignado por la propia herramienta SPSS Modeler, a cada una de las respuestas que hemos filtrado con el

nodo “tipo” (no se observan los IDs de las preguntas, ID cuestionario e ID estudio debido a que se han excluido manualmente con el nodo “tipo”).



Gráfico 12 - Importancia del predictor

La idea del gráfico de la importancia del predictor es centrar los esfuerzos de modelado en las preguntas que importan más, e ignorar las que importan menos. SPSS Modeler ofrece un predictor de importancia que mide la importancia relativa (eje X) de cada pregunta con respecto a las demás. En el Gráfico 12 podemos observar las 18 entradas (columnas de las respuestas) seleccionadas por su importancia relativa en el modelo K-Means, las demás preguntas que no aparecen en el gráfico fueron excluidas por ser consideradas de poca importancia para el modelo (Ver Anexo A, con las preguntas).

Podemos suponer que las preguntas descartadas por el modelo K-Means son de poca importancia debido a que, en los tres clústeres generados, las respuestas dadas por los entrevistados son similares. Para comprobar que, para esas preguntas los usuarios han contestado de manera similar, con independencia del clúster, analizamos la pregunta 1 (excluida por su poca importancia) en el Gráfico 13. Como podemos ver, independientemente del clúster, el área de mayor interés para los ciudadanos es el trabajo y en segundo lugar la sanidad. El orden de interés de las áreas es el mismo para los 3 clústeres.

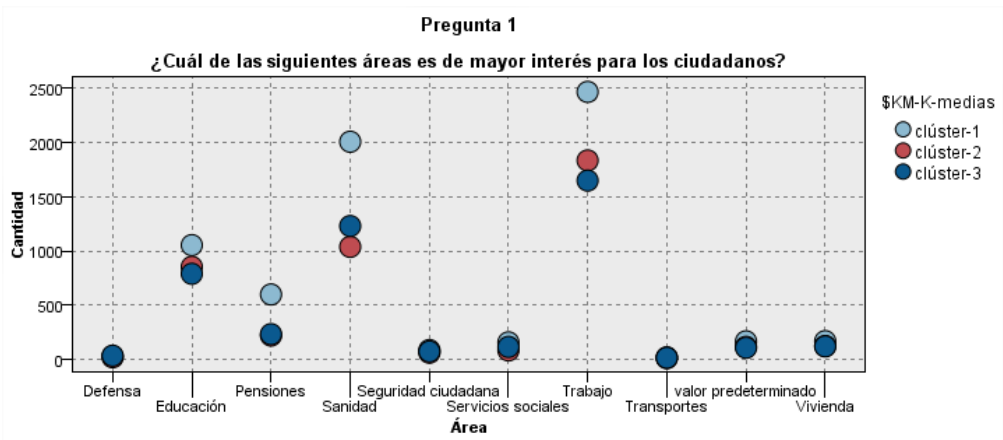


Gráfico 13 - Pregunta 1

En el Gráfico 13 podemos observar que existe una distancia entre los clústeres para algunas de las áreas (los círculos no están solapados). Esto se debe a la cantidad de usuarios asignados a cada clúster, ya que el eje Y mide la cantidad de usuarios que han contestado esa área en concreto. Lo que se pretende analizar con el gráfico es el orden de preferencia de los clústeres, y no la cantidad de usuarios que han respondido en cada clúster.

Por otro lado, de las preguntas clasificadas de mayor importancia para el modelo, analizaremos una en concreto, la pregunta 28. En el Gráfico 14 podemos apreciar una mayor variación de opinión entre clústeres, ya que se puede observar claramente que el clúster 1 ha valorado el grado de confianza y seguridad que transmite el médico de manera superior a los otros dos clústeres, mientras que el clúster 2 parece mantenerse en la media, y el clúster 3 lo ha valorado negativamente. Por tanto, podemos concluir que hay una discrepancia de opinión entre los entrevistados de cada clúster, y que es correcto asignar a esta pregunta una mayor importancia para el modelo.

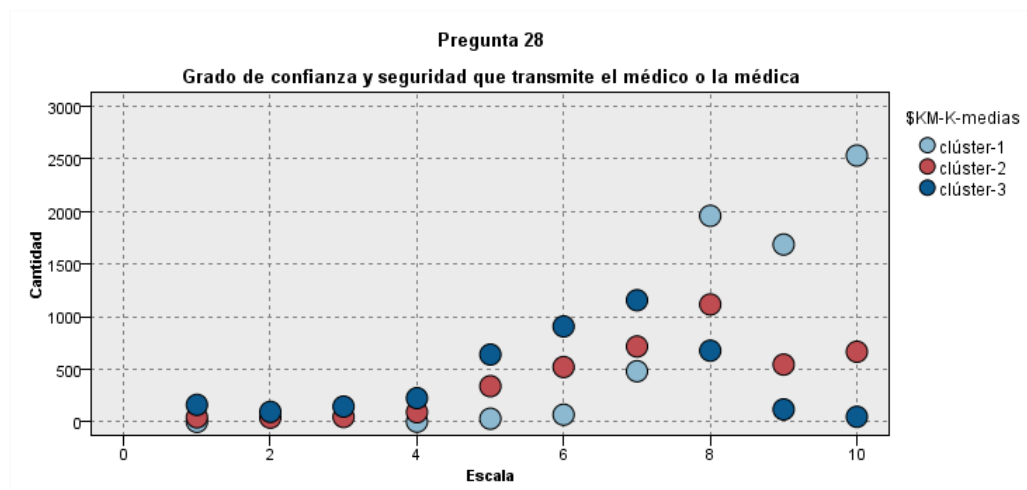


Gráfico 14 - Pregunta 26

Como conclusión, hemos observado que el porcentaje de tamaños de los 3 clústeres identificados por el modelo K-Means se asemeja a los porcentajes estadísticos analizados en la tarea de Exploración de los datos, más en concreto, a los porcentajes de la Tabla 2 - Nota Atención Primaria, por lo que podríamos decir que el modelo K-Means se aproxima a los datos estadísticos analizados en las fases anteriores.

Por otro lado, también podemos comprender mediante los gráficos Gráfico 13 y Gráfico 14, como la herramienta SPSS Modeler asigna la importancia a cada pregunta observada en el Gráfico 12 - Importancia del predictor.

Fase de Evaluación

En esta fase realizaremos una evaluación de los resultados obtenidos en la fase de modelado desde una perspectiva empresarial, comprobando que son válidos y concluyentes con respecto a los objetivos del negocio establecidos al inicio del proyecto y asegurando que se puedan utilizar para identificar y mejorar distintos aspectos de la atención primaria de la sanidad pública española [8].

3.1.18 Evaluación de los resultados

En la fase de comprensión del negocio, identificamos que su principal objetivo es obtener información relevante sobre la satisfacción y opinión de los usuarios residentes en España con respecto a la atención primaria de la sanidad pública, determinar aspectos positivos y negativos de los servicios prestados y plantear posibles mejoras aplicables a los mismos. Con el motivo de alcanzar los objetivos planteados, se ha decidido realizar un proyecto de minería de datos para identificar distintos sectores dentro de la población de residentes en España que opinen de manera similar, y estudiar los distintos factores que les conducen a crear tal opinión.

Para identificar los distintos sectores, se ha optado por realizar un modelado de segmentación. Se han planteado tres modelos distintos: K-Means, Kohonen y Bietápico. Después de realizar pruebas con cada uno de ellos, se ha concluido que el modelo K-Means es el que más se adapta a los objetivos del proyecto y más se aproxima a los datos estadísticos.

Como se ha mostrado en el apartado de evaluación del modelo de la fase de modelado, el algoritmo de K-Means identifica 3 clústeres entre todos los registros recogidos en la fase de selección de datos. Como se puede observar, los tamaños en porcentaje de los clústeres se asemejan a los porcentajes de valoración (del 1 al 10) analizados en la tarea Exploración de los datos de la Fase de comprensión de datos. La herramienta SPSS Modeler permite visualizar los clústeres por separado para ver las preguntas que resultan ser decisivas (dentro de las 18 seleccionadas) para que el registro pertenezca a tal clúster.

Analizando las 18 preguntas seleccionadas por el predictor de importancia (Gráfico 12 - Importancia del predictor), podemos observar que la pregunta 13 parece ser decisiva para que el registro pertenezca al clúster número 2. La pregunta 13 corresponde a la siguiente pregunta del barómetro sanitario:

“Durante los últimos doce meses, ¿ha acudido Ud. a la consulta de un médico o una médica de cabecera (o de familia) para una verdadera consulta, es decir, no para pedir una cita, hacerse una radiografía, un análisis u otra prueba, ni para acompañar a otra persona?”

Las posibles respuestas a la pregunta eran las siguientes:

| idrespues... integer | respuesta character varying (500) |
|-------------------------|--------------------------------------|
| 1 | Sí |
| 2 | No |
| 3 | No recuerda |
| 9 | N.C. |

Tabla 13 - Respuestas pregunta 13

Observando los gráficos en la Ilustración 15, donde el eje X representa la respuesta y el eje Y la frecuencia, podemos comprender el motivo de la importancia de la pregunta para generar el clúster 2, debido a que los usuarios que han contestado “No”, no tuvieron suficiente experiencia como para generar una opinión relevante sobre la atención primaria de la sanidad. Por otro lado, tanto el clúster 1 como el clúster 3 han contestado “Sí” a la

pregunta, por tanto, podemos concluir que la pregunta 13 sólo tiene importancia para identificar el clúster 2, que ha contestado “No”.

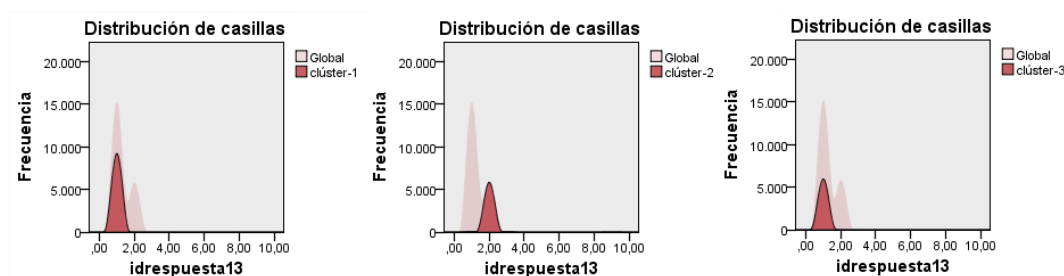


Ilustración 15 - Distribución de casillas (Pregunta 13)

Teniendo en cuenta lo anterior, nos centraremos en los clústeres 1 y 3 debido a que contienen a los usuarios que han acudido a sus médicos de cabecera en los últimos 12 meses y, por tanto, tienen una opinión más reciente y relevante con respecto a la atención primaria.

Analizando las preguntas que la herramienta SPSS Modeler considera de alta importancia para la segmentación de los clústeres (Gráfico 12), podemos identificar una serie de preguntas (preguntas 26 – 34, Tabla 14) que solicitan al usuario valorar entre 1 y 10 distintos aspectos de la asistencia sanitaria pública, referidas a la atención primaria. El enunciado de la pregunta es el siguiente:

“Por su experiencia personal o por la idea que Ud. tenga, quisiera que valorase los siguientes aspectos de la asistencia sanitaria pública, referidos a la atención que se presta en las consultas de médico/a/s de cabecera (o de familia) y pediatría. Utilice para ello una escala de 1 a 10, en la que el 1 significa que lo valora como “totalmente insatisfactorio” y el 10 que lo valora como “totalmente satisfactorio”.”

| Pregunta | Media | | Diferencia |
|---|-----------|-----------|------------|
| | Clúster 1 | Clúster 3 | |
| 26. El trato recibido del personal sanitario | 8,6 | 6,34 | 2,26 |
| 27. El tiempo dedicado por el médico o la médica a cada enfermo o enferma | 8,35 | 5,57 | 2,78 |
| 28. La confianza y seguridad que transmite el médico o la médica | 8,89 | 6,08 | 2,81 |
| 29. El conocimiento del historial y seguimiento de los problemas de salud de cada usuario o usuaria | 8,73 | 5,92 | 2,81 |
| 30. La información recibida sobre su problema de salud | 8,75 | 5,92 | 2,83 |
| 31. Los consejos del médico o de la médica sobre alimentación, ejercicio, tabaco, alcohol, etc. | 8,58 | 5,63 | 2,95 |
| 32. El tiempo que tarda el médico o la médica en verle desde que pide cita | 7,87 | 5,2 | 2,67 |
| 33. El tiempo que se tarda en hacer las pruebas diagnósticas | 6,91 | 4,1 | 2,81 |
| 34. La atención prestada por la/el enfermera/o | 8,45 | 6,19 | 2,26 |

Tabla 14 - Preguntadas destacadas

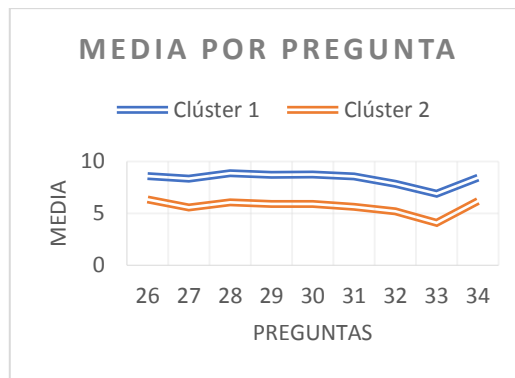


Gráfico 15 - Media por preguntas

Como se puede observar en el Gráfico 15, obtenido con los datos de la Tabla 14, la media de las respuestas dadas por los usuarios asignados al clúster 1 es superior en todos los casos a la media de las respuestas de los usuarios del clúster 3. También podemos observar que la diferencia entre medias es mayor en la pregunta 31, mientras que la diferencia de la media es menor en las preguntas 26 y 34.

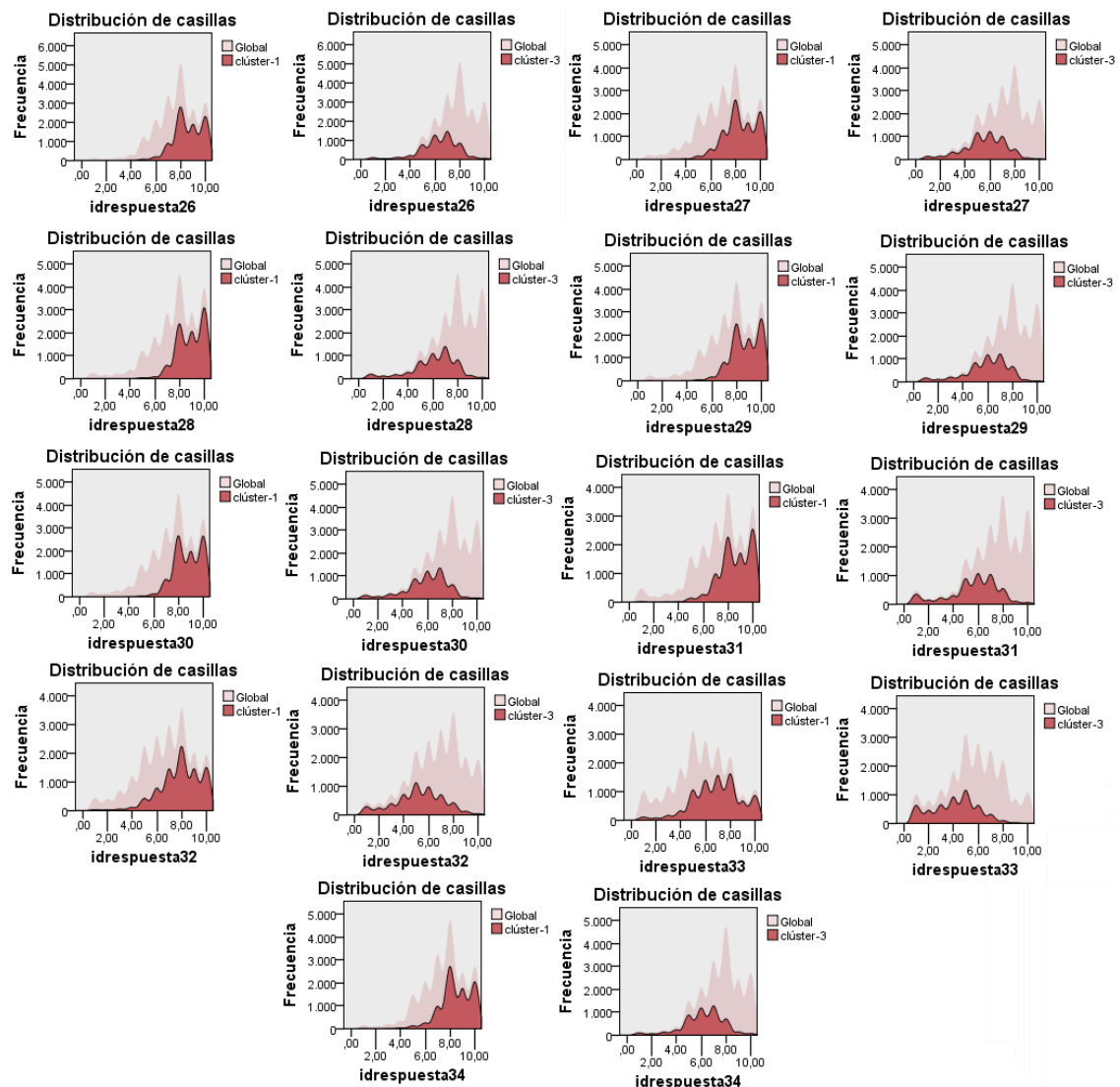


Ilustración 16 - Distribución de casillas (Preguntas 26-34, Clústeres 1 y 3)

En la Ilustración 16 podemos observar por cada pregunta listada en la Tabla 14, dos gráficos, uno para el clúster 1 y otro para el clúster 3 (en total 18 gráficos). Estos gráficos indican la frecuencia (eje Y) de cada respuesta dada por cada uno de los clústeres (eje X, respuestas del 1 al 10).

Con los datos analizados y obtenidos en la Tabla 12 podemos concluir que los resultados expresan con claridad que existe un sector mayoritario (43,61%), representado mediante el clúster 1, que está muy satisfecho con los servicios prestados en la atención primaria de la sanidad pública española. Por otro lado, tenemos dos sectores de menor tamaño que representan a los usuarios poco satisfechos con la atención primaria (28,22%), y los usuarios que no han utilizado los servicios públicos primarios en los últimos 12 meses (28,18%). Estos porcentajes encajan con los obtenidos estadísticamente en la Tabla 2 de la tarea de Exploración de los datos, donde podíamos observar un grupo de usuarios representado por un 40%, que habían acudido a los servicios de la atención primaria y la había valorado positivamente con una nota superior al 7, estos corresponderían al clúster 1. Por otro lado, tenemos otro grupo de usuarios (32%) que también han acudido a los servicios en los últimos 12 meses, pero lo han valorado negativamente con una nota inferior al 7, este grupo correspondería al clúster 3. Por último, tenemos un grupo representado por un 28% de los usuarios, que no han acudido a los servicios de atención primaria en los últimos 12 meses, y por tanto no tienen una opinión válida y relevante para el estudio, estos corresponderían al clúster 2.

También se han realizado descubrimientos relevantes relacionados con los aspectos de mayor importancia para el usuario a la hora de crear una opinión sobre los servicios primarios, como la importancia de la confianza y seguridad que transmite el médico, el tiempo dedicado por el médico a cada enfermo o los consejos del médico sobre alimentación, ejercicio, tabaco, alcohol, etc. El modelo K-Means nos ha mostrado un listado de 18 preguntas relevantes, en el Gráfico 12 - Importancia del predictor, que han sido determinantes para que cada entrevistado haya sido asignado a un clúster (Anexo B).

3.1.19 Proceso de revisión

Los procesos realizados hasta este punto han sido ejecutados según lo previsto con pocas complicaciones, entre ellas, cabe destacar una que se ha detectado a tiempo y se ha tenido en cuenta, relacionada con la construcción de los modelos seleccionados (K-Means, Kohonen y Bietápico), debido a los valores de las respuestas “No sabe/No contesta”. Esto se debe a que, en el barómetro, estas respuestas son codificadas con los valores 98 y 99, por lo que, si no se tiene en cuenta en la construcción del modelo, esto dificulta la visualización de los resultados de los registros de cada clúster.

3.1.20 Determinar próximos pasos

Dado que los resultados se adaptan a los objetivos comerciales de minería de datos, el próximo paso a realizar es la ejecución de la fase de implementación y conclusión del proyecto.

Fase de Implementación

En esta última fase de la metodología CRISP-DM, utilizaremos los conocimientos obtenidos para explicar al cliente como implementar las mejoras en su organización y

también exponer los resultados de manera que el cliente lo pueda entender fácilmente. En esta fase también realizaremos un informe final con las conclusiones del proyecto [8]

3.1.21 Plan de implantación

En esta tarea indicaremos los pasos a realizar por la administración para mejorar la opinión de los usuarios que han valorado negativamente la atención primaria de la sanidad pública. Para ello, recomendamos poner foco en las 18 preguntas recogidas en el Gráfico 12 - Importancia del predictor, que pueden ser visualizadas en el anexo B. Por ejemplo, entre las preguntas recogidas, observamos que el tiempo que se tarda para que el médico pueda atender al usuario o el tiempo que se tarda en hacer las pruebas diagnósticas, son aspectos claves para identificar a los usuarios que no se encuentran satisfechos con los servicios, por lo que se recomendaría tomar medidas para reducir esos tiempos. También destacamos la importancia de la confianza y seguridad que debe transmitir el médico al paciente, ya que es algo que se debe tener en cuenta si se quiere mejorar la opinión de los usuarios.

Para la implantación del proyecto en el negocio, será necesario recoger los próximos barómetros sanitarios realizados en una única base de datos, y también introducir los barómetros sanitarios realizados en los años anteriores, para dar una mejor visión de la evolución de la satisfacción de los usuarios a lo largo de los años.

3.1.22 Plan de monitoreo y mantención

Como el objetivo del proyecto es analizar la opinión actual de los usuarios de la sanidad pública, y para ello se realiza un barómetro sanitario anual, es necesario crear un plan de supervisión y mantenimiento de la implementación del proyecto. Las encuestas son realizadas de manera anual, por lo que los datos tienen que ser analizados, preparados y procesados cada vez que se concluya el año, para formular nuevas acciones a realizar por parte de la administración.

Como plan de monitoreo y mantenimiento se podrían establecer los siguientes procesos:

- Comprobar que las acciones recomendadas en la implantación se están cumpliendo por parte de la administración pública.
- Realizar una comprobación de las preguntas incluidas en las encuestas realizadas en el año actual con respecto a las preguntas de otros años. Y actualizar las tablas de preguntas y respuestas si fuera necesario.
- Una vez recogidas todas las encuestas realizadas, introducirlas en una tabla de base de datos temporal para hacer un ejercicio de análisis, para buscar posibles valores erróneos antes de añadirlas en la tabla principal de hechos (con los demás años).
- Es recomendable pasar a tablas históricas los datos muy antiguos, anteriores a 10 años, para que el volumen de datos y tiempos de proceso no sean demasiado excesivos.
- Por último, realizar un informe con los resultados obtenidos, generando gráficos para una mejor visualización, y destacar nuevas acciones para mejorar la satisfacción de los usuarios.

3.1.23 Informe final

Actualmente, el Ministerio de Sanidad, Servicios Sociales e Igualdad realiza anualmente, junto al Centro de Investigaciones Sociológicas, un barómetro sanitario que pretende conocer de manera general la opinión de los usuarios de la sanidad pública española. A partir de ese barómetro se realizan estadísticas para conocer el porcentaje de usuarios satisfechos con tales servicios, pero se desconocen los detalles de distintos factores claves que hacen que el usuario este contento o no con los servicios prestados. Con este proyecto de minería de datos se pretende encontrar distintos sectores de la población que agrupen a los usuarios que opinen de manera similar entre ellos, para analizar sus respuestas dadas a cada una de las preguntas del barómetro, e identificar los distintos servicios o factores que influyen en su opinión.

Para la realización del proyecto de minería de datos, se han seguido las directrices de la metodología CRISP-DM, una de las más utilizadas en este tipo de proyectos. En la metodología se detallan cada una de las fases a seguir para conseguir los objetivos planteados, de manera ordenada y clara. Dentro de la metodología se recogen y preparan los datos, para posteriormente aplicarles una técnica de modelado y así generar resultados y descubrimientos que den respuesta al problema inicialmente planteado por el negocio.

Después de aplicar tres técnicas de modelado (K-Means, kohonen y bietápico) a los datos previamente seleccionados y preparados, se han analizado por separado cada uno de los resultados obtenidos, para finalmente elegir el modelo K-Means, ya que sus resultados eran los que más se adaptaban al objetivo del proyecto, y el que más información relevante aportaba. Con el modelo K-Means generado, se han podido observar los distintos grupos (clústeres) encontrados dentro de la población de usuarios que contestaron las encuestas del barómetro sanitario, y las preguntas que fueron decisivas para identificarlos en sus clústeres correspondientes. A partir de esas preguntas pudimos identificar los factores en los que hay que poner foco si se quiere aumentar la satisfacción de los usuarios.

Para la implantación del proyecto, se ha recomendado tener en cuenta las preguntas identificadas como importantes en el modelo K-Means, para realizar acciones y medidas que mejoren la satisfacción de los usuarios de la atención primaria. Por otro lado, será necesario crear una base de datos con los resultados del barómetro sanitario de los años anteriores y los resultados de los próximos que se realicen, para descubrir nuevos aspectos a tener en cuenta para seguir mejorando.

3.1.24 Revisión del proyecto

Esta es la actividad final de la metodología CRISP-DM. En este paso evaluaremos los procesos realizados correctamente y los que no, también analizaremos posibles mejoras para los futuros proyectos, para así obtener mejores resultados.

Cabría destacar que en la fase de recolección de datos podríamos haber recogido los resultados de algunos años más, anteriores a 2015, para tener un mayor número de registros y obtener resultados aún más precisos. Por otro lado, en la fase de modelado podríamos haber probado otros tipos de técnicas de modelado, no solo otras técnicas de segmentación sino también técnicas de clasificación, por si estas técnicas nos aportaban información relevante a los resultados.

4 Conclusiones y trabajo futuro

Conclusiones

Actualmente, para la implementación de cualquier proyecto tecnológico es altamente recomendable seguir las directrices de una metodología, ya que son creadas y definidas a partir de la experiencia, y te indican paso a paso las acciones a realizar en el proyecto para obtener los mejores resultados y alcanzar tus objetivos.

Existe un gran número de metodologías disponibles para cada tipo de tecnología que se pueden acceder fácilmente a través de Internet y de bibliografía especializada. Una metodología de las más comúnmente utilizadas para proyectos de minería de datos es CRISP-DM. Esta metodología describe las fases normales de un proyecto, las tareas de cada fase y una explicación de las relaciones entre cada una de ellas.

La metodología CRISP-DM ha sido de gran ayuda para el desarrollo del proyecto de análisis de datos sanitarios, ya que se trataba de un proyecto de minería de datos para el que se conocían los objetivos iniciales, pero no las actividades a realizar, los pasos a seguir, ni las técnicas de modelado que podrían dar resultados relevantes que ayudasen a alcanzar los objetivos planteados.

Por otro lado, en un proyecto de minería una de las fases más importantes es el modelado, existiendo modelos para diferentes tareas, tales como la clasificación y la segmentación. Teniendo en cuenta que se disponía de los datos sanitarios completos, sin aportación de nueva información al comienzo del estudio se consideró que la tarea de minería más apropiada era la segmentación, que tiene como principal objetivo detectar segmentos o clústeres en los datos para conocer su disposición actual. Elegida la tarea se plantearon tres posibles técnicas o algoritmos para resolverla, resultando el método K-Means el que mejores resultados ofrecía, teniendo en cuenta los datos estadísticos previos.

Tras aplicar la metodología CRISP-DM y realizar un análisis detallado de los resultados obtenidos con el modelado de datos sanitarios, se pueden identificar tres grandes sectores entre los residentes de España, el de mayor tamaño corresponde a los usuarios que están satisfechos o muy satisfechos con los servicios prestados en la atención primaria. Con la segmentación realizada podemos concluir que la población, en su gran mayoría, está satisfecha con la atención primaria de la sanidad pública, aunque existen algunos factores que se deben tener en cuenta para formular acciones y medidas, y así mejorar y aumentar el número de usuarios satisfechos.

Trabajo futuro

En futuros proyectos a corto plazo, se podrían analizar los otros tres sectores de la sanidad pública española que no se incluyeron en el alcance de este proyecto, la atención especializada, la hospitalización y los servicios de urgencias. Con el análisis de los cuatro sectores, se podría visualizar el nivel de satisfacción de los usuarios de manera global en relación a la totalidad de los servicios prestados por la sanidad pública española. Otro proyecto a futuro de gran interés es, aplicando la metodología CRISP-DM, realizar un proyecto de minería de datos de predicción de la satisfacción de los futuros usuarios de la sanidad pública española, a partir de las encuestas del barómetro sanitario.

Referencias

- [1] *Aníbal goicochea - CRISP-DM*. (05 de 06 de 2017). Obtenido de Tecnologías de la Información y Estrategia: <https://anibalgoicochea.com/2009/08/11/crisp-dm-una-metodologia-para-proyectos-de-mineria-de-datos/>
- [2] *Banco de datos, Ministerio de Sanidad, Servicios Sociales e Igualdad*. (05 de 06 de 2017). Obtenido de <http://www.msssi.gob.es/estadEstudios/estadisticas/bancoDatos.htm>
- [3] *Barómetro Sanitario, Microdatos*. (05 de 06 de 2017). Obtenido de <http://www.msssi.gob.es/estadisticas/microdatos.do>
- [4] *Barómetro Sanitario, Notas Técnicas*. (05 de 06 de 2017). Obtenido de <http://www.msssi.gob.es/estadEstudios/estadisticas/NotasTecnicas/NotasTecnicasBS3.pdf>
- [5] *CRISP-DM - Guide*. (05 de 06 de 2017). Obtenido de <https://www.the-modeling-agency.com/crisp-dm.pdf>
- [6] *Grupo de Meteorología de Santander*. (05 de 06 de 2017). Obtenido de Minería de Datos. Redes Bayesianas y Neuronales: http://www.meteo.unican.es/es/research/mineria_datos
- [7] *IBM Knowledge Center*. (05 de 06 de 2017). Obtenido de Modelos de conglomerado: https://www.ibm.com/support/knowledgecenter/es/SS3RA7_15.0.0/com.ibm.spss.modeler.help/nodes_clusteringmodels.htm
- [8] *Manual CRISP-DM de IBM SPSS Modeler*. (05 de 06 de 2017). Obtenido de <ftp://public.dhe.ibm.com/software/analytics/spss/documentation/modeler/15.0/es/CRISP-DM.pdf>
- [9] *Metodología para el desarrollo de proyectos en Minería de Datos CRISP-DM*. (05 de 06 de 2017). Obtenido de http://www.oldemarrodriguez.com/yahoo_site_admin/assets/docs/Documento_CRI SP-DM.2385037.pdf
- [10] *Microsoft - Conceptos de minería de datos*. (05 de 06 de 2017). Obtenido de <https://msdn.microsoft.com/es-es/library/ms174949.aspx>
- [11] *Wikipedia - Algoritmo*. (05 de 06 de 2017). Obtenido de <https://es.wikipedia.org/wiki/Algoritmo>
- [12] *Wikipedia - Algoritmo de agrupamiento*. (05 de 06 de 2017). Obtenido de Clustering: https://es.wikipedia.org/wiki/Algoritmo_de_agrupamiento
- [13] *Wikipedia - Minería de datos*. (05 de 06 de 2017). Obtenido de https://es.wikipedia.org/wiki/Miner%C3%ADa_de_datos
- [14] *Diseño Data Warehouse: Hechos Y Dimensiones; Modelo Estrella Vs Copo De Nieve*. Obtenido de <http://mundodb.es/diseño-data-warehouse-hechos-y-dimensiones-modelo-estrella-vs-copo-de-nieve>
- [15] *Pruebas de caja negra*. Obtenido de Testing funcional a nivel de datos externos: <https://www.globetesting.com/2012/08/pruebas-de-caja-negra/>

Glosario

| | |
|---------------------|--|
| Algoritmo | Conjunto prescrito de instrucciones o reglas bien definidas, ordenadas y finitas que permite llevar a cabo una actividad mediante pasos sucesivos que no generen dudas a quien deba hacer dicha actividad. |
| Backup | Copia de seguridad, copia de respaldo |
| Barómetro Sanitario | Estudio de opinión que viene realizándose con una periodicidad anual desde 1993 por el Ministerio de Sanidad, Servicios Sociales e Igualdad en colaboración con el Centro de Investigaciones Sociológicas. |
| Bietápico | Se utiliza para agrupar el conjunto de datos en grupos distintos cuando se desconoce su significado, intenta revelar patrones en el conjunto de campos de entrada. |
| Clustering | Procedimiento de agrupación de una serie de vectores de acuerdo con un criterio. Esos criterios son por lo general distancia o similitud. |
| CRIPS-DM | Cross Industry Standard Process for Data Mining, guías de referencia utilizada en proyectos de minería de datos. |
| BIC | Criterio bayesiano de Schwarz - Criterio para la selección de modelos entre un conjunto finito de modelos. |
| K-Means | Se utiliza para agrupar conjuntos de datos en distintos clústeres que en un principio se desconocen. Este modelo intenta revelar los patrones en el conjunto de datos, agrupando las instancias de manera que las de un mismo clúster tienden a ser similares entre ellas. |
| Kohonen | Los mapas de Kohonen son mapas auto-organizados que, son entrenados usando un algoritmo de aprendizaje no supervisado y generan una representación discreta del espacio de las muestras de entrada. |
| Minería de datos | Campo de la estadística y de la informática que se encarga de procesar, analizar y descubrir patrones e información de manera automática o semi-automática en grandes volúmenes de datos. |
| pgAdmin | Herramienta de administración de bases de datos PostgreSQL. |
| Red Neuronal | Modelos no lineales, inspirados en el funcionamiento del cerebro, que fueron diseñados para resolver una gran variedad de problemas. |

SPSS Modeler

Plataforma de análisis predictiva diseñada para realizar predicciones sobre decisiones llevadas a cabo por personas, sistemas o empresas.

Anexos

A Preguntas seleccionadas

| ID | Pregunta |
|----|--|
| 1 | Para empezar, ¿podría decirme cuál de las siguientes áreas que le voy a leer es la que considera Ud. de mayor interés para los ciudadanos y las ciudadanas? |
| 2 | Vamos a centrarnos ahora en el tema de la sanidad. De las siguientes afirmaciones que aparecen en esta tarjeta, ¿cuál expresa mejor su opinión sobre el sistema sanitario en nuestro país? |
| 3 | ¿Está Ud. satisfecho/a o insatisfecho/a con el modo en que el sistema sanitario público funciona en España? Para contestarme, utilice esta tarjeta en la que el 1 significa que está Ud. Muy insatisfecho/a y el 10 que está muy satisfecho/a. |
| 4 | En su opinión, ¿la sanidad pública está mejor gestionada si lo hace...? |
| 5 | Si Ud. o algún miembro de su hogar tuvieran que utilizar un servicio sanitario y Ud. pudiera elegir, ¿acudiría a un centro público o privado, cuando se tratara de consultas de atención primaria (consultas de médicos/as de cabecera (o de familia) y pediatría)? |
| 6 | Si Ud. o algún miembro de su hogar tuvieran que utilizar un servicio sanitario y Ud. pudiera elegir, ¿acudiría a un centro público o privado, cuando se tratara de consultas de asistencia especializada (consultas de especialistas, salvo dentistas)? |
| 7 | Si Ud. o algún miembro de su hogar tuvieran que utilizar un servicio sanitario y Ud. pudiera elegir, ¿acudiría a un centro público o privado, cuando se tratara de ingreso en hospital? |
| 8 | Si Ud. o algún miembro de su hogar tuvieran que utilizar un servicio sanitario y Ud. pudiera elegir, ¿acudiría a un centro público o privado, cuando se tratara de urgencias? |
| 9 | Por su propia experiencia o por la idea que Ud. tenga, me gustaría que valorase los siguientes servicios sanitarios públicos. Para contestar, utilice por favor una escala de 1 a 10, en el que el 1 significa que le parece “totalmente insatisfactorio” y el 10 “totalmente satisfactorio.”: Consultas de atención primaria (consultas de médicos/as de cabecera (o de familia) y pediatría) en centros de salud |
| 10 | Por su propia experiencia o por la idea que Ud. tenga, me gustaría que valorase los siguientes servicios sanitarios públicos. Para contestar, utilice por favor una escala de 1 a 10, en el que el 1 significa que le parece “totalmente insatisfactorio” y el 10 “totalmente satisfactorio.”: Consultas de atención especializada (consultas de especialistas en centros públicos) |
| 11 | Por su propia experiencia o por la idea que Ud. tenga, me gustaría que valorase los siguientes servicios sanitarios públicos. Para contestar, utilice por favor una escala de 1 a 10, en el que el 1 significa que le parece “totalmente insatisfactorio” y el 10 “totalmente satisfactorio.”: Urgencias en hospitales públicos |
| 12 | Por su propia experiencia o por la idea que Ud. tenga, me gustaría que valorase los siguientes servicios sanitarios públicos. Para contestar, utilice por favor una escala de 1 a 10, en el que el 1 significa que le parece “totalmente insatisfactorio” y el 10 “totalmente satisfactorio.”: Ingreso y asistencia en hospitales públicos |

| | |
|----|---|
| 13 | Durante los últimos doce meses, ¿ha acudido Ud. a la consulta de un médico o una médica de cabecera (o de familia) para una verdadera consulta, es decir, no para pedir una cita, hacerse una radiografía, un análisis u otra prueba, ni para acompañar a otra persona? |
| 14 | Y durante estos últimos doce meses, ¿puede recordar cuántas veces acudió a un médico o a una médica de cabecera (o de familia) de la sanidad pública? |
| 15 | Y durante estos últimos doce meses, ¿puede recordar cuántas veces acudió a un médico o a una médica de cabecera (o de familia) de la sanidad privada? |
| 16 | Y en la última visita al/a la médico/a cabecera (o de familia) de la sanidad pública ¿Pudo participar en las decisiones sobre su problema de salud y su tratamiento, tanto como Ud. Hubiera deseado? |
| 17 | Y en la última visita al/a la médico/a cabecera (o de familia) de la sanidad pública ¿Le dio el médico o la médica la oportunidad de hacer preguntas o plantear preocupaciones? |
| 18 | Y en la última visita al/a la médico/a cabecera (o de familia) de la sanidad pública ¿Tuvo Ud. la sensación de que el médico o la médica le prestaba menos atención de lo que a Ud. le hubiera gustado, por estar pendiente del ordenador? |
| 19 | Y en la última visita al/a la médico/a cabecera (o de familia) de la sanidad pública ¿Tuvo Ud. que esperar más de una hora en la sala de espera hasta entrar en la consulta? |
| 20 | En las visitas de los últimos doce meses a algún/a médico/a de cabecera (o de familia) de la sanidad pública, ¿ha sufrido algún tipo de error en la asistencia sanitaria recibida? |
| 21 | ¿Y este error afectó de una manera muy grave, bastante grave, poco o nada grave a su salud? |
| 22 | En general, la atención que ha recibido en las consultas del médico o de la médica de cabecera (o de familia) de la sanidad pública ha sido |
| 23 | Y respecto de lo que Ud. esperaba, esta atención ha sido |
| 24 | Cuando pide cita con el médico o la médica de su centro de salud de la sanidad pública para ir en el mismo día, se la dan... |
| 25 | La última vez que pidió cita al/a la médico/a de cabecera, ¿con cuántos días se la dieron? (ENTREVISTADOR/A: Si la persona entrevistada contesta un periodo de días, recoger el número días más alto). |
| 26 | Por su experiencia personal o por la idea que Ud. tenga, quisiera que valorase los siguientes aspectos de la asistencia sanitaria pública, referidos a la atención que se presta en las consultas de médico/a/s de cabecera (o de familia) y pediatría. Utilice para ello una escala de 1 a 10, en la que el 1 significa que lo valora como “totalmente insatisfactorio” y el 10 que lo valora como “totalmente satisfactorio”: El trato recibido del personal sanitario |
| 27 | Por su experiencia personal o por la idea que Ud. tenga, quisiera que valorase los siguientes aspectos de la asistencia sanitaria pública, referidos a la atención que se presta en las consultas de médico/a/s de cabecera (o de familia) y pediatría. Utilice para ello una escala de 1 a 10, en la que el 1 significa que lo valora como “totalmente insatisfactorio” y el 10 que lo valora como “totalmente satisfactorio”: El tiempo dedicado por el médico o la médica a cada enfermo o enferma |

| | |
|----|---|
| 28 | Por su experiencia personal o por la idea que Ud. tenga, quisiera que valorase los siguientes aspectos de la asistencia sanitaria pública, referidos a la atención que se presta en las consultas de médico/a/s de cabecera (o de familia) y pediatría. Utilice para ello una escala de 1 a 10, en la que el 1 significa que lo valora como “totalmente insatisfactorio” y el 10 que lo valora como “totalmente satisfactorio”: La confianza y seguridad que transmite el médico o la médica |
| 29 | Por su experiencia personal o por la idea que Ud. tenga, quisiera que valorase los siguientes aspectos de la asistencia sanitaria pública, referidos a la atención que se presta en las consultas de médico/a/s de cabecera (o de familia) y pediatría. Utilice para ello una escala de 1 a 10, en la que el 1 significa que lo valora como “totalmente insatisfactorio” y el 10 que lo valora como “totalmente satisfactorio”: El conocimiento del historial y seguimiento de los problemas de salud de cada usuario o usuaria |
| 30 | Por su experiencia personal o por la idea que Ud. tenga, quisiera que valorase los siguientes aspectos de la asistencia sanitaria pública, referidos a la atención que se presta en las consultas de médico/a/s de cabecera (o de familia) y pediatría. Utilice para ello una escala de 1 a 10, en la que el 1 significa que lo valora como “totalmente insatisfactorio” y el 10 que lo valora como “totalmente satisfactorio”: La información recibida sobre su problema de salud |
| 31 | Por su experiencia personal o por la idea que Ud. tenga, quisiera que valorase los siguientes aspectos de la asistencia sanitaria pública, referidos a la atención que se presta en las consultas de médico/a/s de cabecera (o de familia) y pediatría. Utilice para ello una escala de 1 a 10, en la que el 1 significa que lo valora como “totalmente insatisfactorio” y el 10 que lo valora como “totalmente satisfactorio”: Los consejos del médico o de la médica sobre alimentación, ejercicio, tabaco, alcohol, etc. |
| 32 | Por su experiencia personal o por la idea que Ud. tenga, quisiera que valorase los siguientes aspectos de la asistencia sanitaria pública, referidos a la atención que se presta en las consultas de médico/a/s de cabecera (o de familia) y pediatría. Utilice para ello una escala de 1 a 10, en la que el 1 significa que lo valora como “totalmente insatisfactorio” y el 10 que lo valora como “totalmente satisfactorio”: El tiempo que tarda el médico o la médica en verle desde que pide la cita |
| 33 | Por su experiencia personal o por la idea que Ud. tenga, quisiera que valorase los siguientes aspectos de la asistencia sanitaria pública, referidos a la atención que se presta en las consultas de médico/a/s de cabecera (o de familia) y pediatría. Utilice para ello una escala de 1 a 10, en la que el 1 significa que lo valora como “totalmente insatisfactorio” y el 10 que lo valora como “totalmente satisfactorio”: El tiempo que se tarda en hacer las pruebas diagnósticas |
| 34 | Por su experiencia personal o por la idea que Ud. tenga, quisiera que valorase los siguientes aspectos de la asistencia sanitaria pública, referidos a la atención que se presta en las consultas de médico/a/s de cabecera (o de familia) y pediatría. Utilice para ello una escala de 1 a 10, en la que el 1 significa que lo valora como “totalmente insatisfactorio” y el 10 que lo valora como “totalmente satisfactorio”: La atención prestada por la/el enfermero/a |
| 35 | En general, ¿cree Ud. que, durante los últimos doce meses, el problema de las listas de espera...? |
| 36 | En su opinión, ¿la atención primaria de la sanidad pública ha mejorado, ha empeorado o sigue igual respecto a los últimos cinco años? |
| 37 | Sexo: |
| 38 | ¿Cuántos años cumplió Ud. en su último cumpleaños? |

| | |
|----|--|
| 39 | ¿Ha ido Ud. a la escuela o cursado algún tipo de estudios? (ENTREVISTADOR/A: en caso negativo, preguntar si sabe leer y escribir). |
| 40 | ¿Cómo describiría Ud. su estado de salud en general: muy bueno, bueno, regular, malo o muy malo? |
| 41 | ¿Tiene Ud. alguna enfermedad o problema de salud crónico? |
| 42 | ¿En cuál de las siguientes situaciones se encuentra Ud. actualmente? |
| 43 | ¿Trabaja/ba Ud. (o la persona que aporta más ingresos al hogar) en la Administración Pública, en una empresa pública, en una empresa privada, en una organización privada sin fines de lucro o en el servicio doméstico? |
| 44 | Con respecto a la nacionalidad, ¿tiene Ud. ...? |

B Resumen Clústeres (Preguntas y Respuestas)

| ID | Pregunta | Clúster 1 | Clúster 2 | Clúster 3 |
|----|--|--------------|--------------|-------------------|
| 9 | Por su propia experiencia o por la idea que Ud. tenga, me gustaría que valorase los siguientes servicios sanitarios públicos. Para contestar, utilice por favor una escala de 1 a 10, en el que el 1 significa que le parece “totalmente insatisfactorio” y el 10 “totalmente satisfactorio.”: Consultas de atención primaria (consultas de médicos/as de cabecera (o de familia) y pediatría) en centros de salud | 8,282 | 7,134 | 5,953 |
| 13 | Durante los últimos doce meses, ¿ha acudido Ud. a la consulta de un médico o una médica de cabecera (o de familia) para una verdadera consulta, es decir, no para pedir una cita, hacerse una radiografía, un análisis u otra prueba, ni para acompañar a otra persona? | Sí | No | Sí |
| 14 | Y durante estos últimos doce meses, ¿puede recordar cuántas veces acudió a un médico o a una médica de cabecera (o de familia) de la sanidad pública? | 4 | Ninguna | 3 |
| 15 | Y durante estos últimos doce meses, ¿puede recordar cuántas veces acudió a un médico o a una médica de cabecera (o de familia) de la sanidad privada? | 0 | Ninguna | 0 |
| 17 | Y en la última visita al/a la médico/a cabecera (o de familia) de la sanidad pública ¿Le dio el médico o la médica la oportunidad de hacer preguntas o plantear preocupaciones? | Sí | N.C. | No |
| 19 | Y en la última visita al/a la médico/a cabecera (o de familia) de la sanidad pública ¿Tuvo Ud. que esperar más de una hora en la sala de espera hasta entrar en la consulta? | No | N.C. | No |
| 20 | En las visitas de los últimos doce meses a algún/a médico/a de cabecera (o de familia) de la sanidad pública, ¿ha sufrido algún tipo de error en la asistencia sanitaria recibida? | No | N.C. | No |
| 22 | En general, la atención que ha recibido en las consultas del médico o de la médica de cabecera (o de familia) de la sanidad pública ha sido | Muy buena | N.C. | Regular |
| 23 | Y respecto de lo que Ud. esperaba, esta atención ha sido | Mejor | 9,000 | Más o menos igual |

| | | | | |
|----|---|-------|-------|-------|
| 26 | Por su experiencia personal o por la idea que Ud. tenga, quisiera que valorase los siguientes aspectos de la asistencia sanitaria pública, referidos a la atención que se presta en las consultas de médico/a/s de cabecera (o de familia) y pediatría. Utilice para ello una escala de 1 a 10, en la que el 1 significa que lo valora como “totalmente insatisfactorio” y el 10 que lo valora como “totalmente satisfactorio”: El trato recibido del personal sanitario | 8,595 | 7,503 | 6,341 |
| 27 | Por su experiencia personal o por la idea que Ud. tenga, quisiera que valorase los siguientes aspectos de la asistencia sanitaria pública, referidos a la atención que se presta en las consultas de médico/a/s de cabecera (o de familia) y pediatría. Utilice para ello una escala de 1 a 10, en la que el 1 significa que lo valora como “totalmente insatisfactorio” y el 10 que lo valora como “totalmente satisfactorio”: El tiempo dedicado por el médico o la médica a cada enfermo o enferma | 8,350 | 6,868 | 5,566 |
| 28 | Por su experiencia personal o por la idea que Ud. tenga, quisiera que valorase los siguientes aspectos de la asistencia sanitaria pública, referidos a la atención que se presta en las consultas de médico/a/s de cabecera (o de familia) y pediatría. Utilice para ello una escala de 1 a 10, en la que el 1 significa que lo valora como “totalmente insatisfactorio” y el 10 que lo valora como “totalmente satisfactorio”: La confianza y seguridad que transmite el médico o la médica | 8,891 | 7,503 | 6,082 |
| 29 | Por su experiencia personal o por la idea que Ud. tenga, quisiera que valorase los siguientes aspectos de la asistencia sanitaria pública, referidos a la atención que se presta en las consultas de médico/a/s de cabecera (o de familia) y pediatría. Utilice para ello una escala de 1 a 10, en la que el 1 significa que lo valora como “totalmente insatisfactorio” y el 10 que lo valora como “totalmente satisfactorio”: El conocimiento del historial y seguimiento de los problemas de salud de cada usuario o usuaria | 8,726 | 7,343 | 5,925 |
| 30 | Por su experiencia personal o por la idea que Ud. tenga, quisiera que valorase los siguientes aspectos de la asistencia sanitaria pública, referidos a la atención que se presta en las consultas de médico/a/s de cabecera (o de familia) y pediatría. Utilice para ello una escala de 1 a 10, en la que el 1 significa que lo valora como “totalmente insatisfactorio” y el 10 que lo valora como “totalmente satisfactorio”: La información recibida sobre su problema de salud | 8,746 | 7,320 | 5,925 |

| | | | | |
|----|--|-------|-------|-------|
| 31 | <p>Por su experiencia personal o por la idea que Ud. tenga, quisiera que valorase los siguientes aspectos de la asistencia sanitaria pública, referidos a la atención que se presta en las consultas de médico/a/s de cabecera (o de familia) y pediatría. Utilice para ello una escala de 1 a 10, en la que el 1 significa que lo valora como “totalmente insatisfactorio” y el 10 que lo valora como “totalmente satisfactorio”: Los consejos del médico o de la médica sobre alimentación, ejercicio, tabaco, alcohol, etc.</p> | 8,579 | 7,144 | 5,626 |
| 32 | <p>Por su experiencia personal o por la idea que Ud. tenga, quisiera que valorase los siguientes aspectos de la asistencia sanitaria pública, referidos a la atención que se presta en las consultas de médico/a/s de cabecera (o de familia) y pediatría. Utilice para ello una escala de 1 a 10, en la que el 1 significa que lo valora como “totalmente insatisfactorio” y el 10 que lo valora como “totalmente satisfactorio”: El tiempo que tarda el médico o la médica en verle desde que pide la cita</p> | 7,871 | 6,435 | 5,202 |
| 33 | <p>Por su experiencia personal o por la idea que Ud. tenga, quisiera que valorase los siguientes aspectos de la asistencia sanitaria pública, referidos a la atención que se presta en las consultas de médico/a/s de cabecera (o de familia) y pediatría. Utilice para ello una escala de 1 a 10, en la que el 1 significa que lo valora como “totalmente insatisfactorio” y el 10 que lo valora como “totalmente satisfactorio”: El tiempo que se tarda en hacer las pruebas diagnósticas</p> | 6,912 | 5,469 | 4,095 |
| 34 | <p>Por su experiencia personal o por la idea que Ud. tenga, quisiera que valorase los siguientes aspectos de la asistencia sanitaria pública, referidos a la atención que se presta en las consultas de médico/a/s de cabecera (o de familia) y pediatría. Utilice para ello una escala de 1 a 10, en la que el 1 significa que lo valora como “totalmente insatisfactorio” y el 10 que lo valora como “totalmente satisfactorio”: La atención prestada por la/el enfermero/a</p> | 8,452 | 7,307 | 6,189 |